

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/013227 A2

(51) International Patent Classification⁷: **A01H**

(21) International Application Number: PCT/US02/25805

(22) International Filing Date: 9 August 2002 (09.08.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/310,847 9 August 2001 (09.08.2001) US
60/336,049 19 November 2001 (19.11.2001) US
60/338,692 11 December 2001 (11.12.2001) US
10/171,468 14 June 2002 (14.06.2002) US

(71) Applicant (for all designated States except US):
MENDEL BIOTECHNOLOGY, INC. [US/US]; 21375
Cabot Boulevard, Hayward, CA 94945 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): **RATCLIFFE, Oliver** [GB/US]; 814 East 21st Street, Oakland, CA 94606 (US). **RIECHMANN, Jose, Luis** [ES/US]; 518 S. El Molino Avenue, #308, Pasadena, CA 91101 (US). **ADAM, Luc, J.** [CA/US]; 25800 Industrial Boulevard, Apt. L403, Hayward, CA 94545 (US). **DUBELL, Arnold, T.** [US/US]; 14857 Wake Avenue, San Leandro, CA 94578 (US). **HEARD, Jacqueline, E.** [US/US]; 810 Guilford Avenue, San Mateo, CA 94402 (US). **PILGRIM, Marsha, L.** [US/US]; 1368 Patrick Henry Drive, Phoenixville, PA 19460 (US). **JIANG, Cai-Zhong** [US/US]; 34495 Heathrow Terrace, Fremont, CA 94555 (US). **REUBER, T., Lynne** [US/US]; 1115 S. Grant Street, San Mateo, CA

94402 (US). **CREELMAN, Robert, A.** [US/US]; 2801 Jennifer Drive, Castro Valley, CA 94546 (US). **PINEDA, Omaira** [CO/US]; 4060 9th Place, Vero Beach, FL 32960 (US). **YU, Guo-Liang** [US/US]; 242 Gravett Drive, Berkeley, CA 94705-1531 (US). **BROUN, Pierre, E.** [FR/US]; 921 Sunnybrae Boulevard, San Mateo, CA 94402 (US).

(74) Agents: **WARD, Michael, R.** et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

(57) Abstract: The invention relates to plant transcription factor polypeptides, polynucleotides that encode them, homologs from a variety of plant species, and methods of using the polynucleotides and polypeptides to produce transgenic plants having advantageous properties compared to a reference plant. Sequence information related to these polynucleotides and polypeptides can also be used in bioinformatic search methods and is also disclosed.

WO 03/013227 A2

YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

This application claims the benefit of US Provisional Application No. 60/310,847, filed August 9, 2001, US Provisional Application No. 60/336,049, filed December 5, 2001, US Provisional Application No. 60/338,692, filed December 11, 2001, and US Non-provisional Application No. 10/171,468, filed June 14, 2002, the entire contents of which are hereby incorporated by reference.

FIELD OF THE INVENTION

This invention relates to the field of plant biology. More particularly, the present invention pertains to compositions and methods for phenotypically modifying a plant.

INTRODUCTION

A plant's traits, such as its biochemical, developmental, or phenotypic characteristics, may be controlled through a number of cellular processes. One important way to manipulate that control is through transcription factors - proteins that influence the expression of a particular gene or sets of genes. Transformed and transgenic plants that comprise cells having altered levels of at least one selected transcription factor, for example, possess advantageous or desirable traits. Strategies for manipulating traits by altering a plant cell's transcription factor content can therefore result in plants and crops with commercially valuable properties. Applicants have identified polynucleotides encoding transcription factors, developed numerous transgenic plants using these polynucleotides, and have analyzed the plants for a variety of important traits. In so doing, applicants have identified important polynucleotide and polypeptide sequences for producing commercially valuable plants and crops as well as the methods for making them and using them. Other aspects and embodiments of the invention are described below and can be derived from the teachings of this disclosure as a whole.

BACKGROUND OF THE INVENTION

Transcription factors (TFs) can modulate gene expression, either increasing or decreasing (inducing or repressing) the rate of transcription. This modulation results in differential levels of gene expression at various developmental stages, in different

tissues and cell types, and in response to different exogenous (e.g., environmental) and endogenous stimuli throughout the life cycle of the organism.

Because transcription factors are key controlling elements of biological pathways, altering the expression levels of one or more transcription factors can change entire biological pathways in an organism. For example, manipulation of the levels of selected transcription factors may result in increased expression of economically useful proteins or metabolic chemicals in plants or to improve other agriculturally relevant characteristics. Conversely, blocked or reduced expression of a transcription factor may reduce biosynthesis of unwanted compounds or remove an undesirable trait. Therefore, manipulating transcription factor levels in a plant offers tremendous potential in agricultural biotechnology for modifying a plant's traits.

The present invention provides novel transcription factors useful for modifying a plant's phenotype in desirable ways.

SUMMARY OF THE INVENTION

In a first aspect, the invention relates to a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide comprising a polypeptide sequence selected from those of the Sequence Listing, SEQ ID NOs:2 to 2N, where $N = 2-561$, or those listed in Table 4, or a complementary nucleotide sequence thereof; (b) a nucleotide sequence encoding a polypeptide comprising a variant of a polypeptide of (a) having one or more, or between 1 and about 5, or between 1 and about 10, or between 1 and about 30, conservative amino acid substitutions; (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID NOs:1 to $(2N - 1)$, where $N = 2-561$, or those included in Table 4, or a complementary nucleotide sequence thereof; (d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c); (e) a nucleotide sequence which hybridizes under stringent conditions over substantially the entire length of a nucleotide sequence of one or more of: (a), (b), (c), or (d); (f) a nucleotide sequence comprising at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides of a sequence of any of (a)-(e), or at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides outside of a region encoding a conserved domain of any of (a)-(e); (g) a

nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide having a biological activity that modifies a plant's characteristic, functions as a transcription factor, or alters the level of transcription of a gene or transgene in a cell; (h) a nucleotide sequence having at least 31% sequence identity to a nucleotide sequence of any of (a)-(g); (i) a nucleotide sequence having at least 60%, or at least 70 %, or at least 80 %, or at least 90 %, or at least 95 % sequence identity to a nucleotide sequence of any of (a)-(g) or a 10 or 15 nucleotide, or at least about 20, or at least about 30 nucleotide region of a sequence of (a)-(g) that is outside of a region encoding a conserved domain; (j) a nucleotide sequence that encodes a polypeptide having at least 31% sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; (k) a nucleotide sequence which encodes a polypeptide having at least 60%, or at least 70 %, or at least 80%, or at least 90 %, or at least 95 % sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; and (l) a nucleotide sequence that encodes a conserved domain of a polypeptide having at least 85%, or at least 90%, or at least 95%, or at least 98% sequence identity to a conserved domain of a polypeptide listed in Table 4, or the Sequence Listing. The recombinant polynucleotide may further comprise a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence. The invention also relates to compositions comprising at least two of the above-described polynucleotides.

In a second aspect, the invention comprises an isolated or recombinant polypeptide comprising a subsequence of at least about 10, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids encoded by the recombinant or isolated polynucleotide described above, or comprising a subsequence of at least about 8, or at least about 12, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids outside a conserved domain.

In a third aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fourth aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fifth aspect, the invention comprises an isolated polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a sixth aspect, the invention comprises an isolated polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

The present invention also encompasses transcription factor variants. A preferred transcription factor variant is one having at least 40% amino acid sequence identity, a more preferred transcription factor variant is one having at least 50% amino acid sequence identity and a most preferred transcription factor variant is one having at least 65% amino acid sequence identity to the transcription factor amino acid sequence SEQ ID NOs:2 to 2N, where $N = 2-561$, and which contains at least one functional or structural characteristic of the transcription factor amino acid sequence. Sequences having lesser degrees of identity but comparable biological activity are considered to be equivalents.

In another aspect, the invention is a transgenic plant comprising one or more of the above-described isolated or recombinant polynucleotides. In yet another aspect, the invention is a plant with altered expression levels of a polynucleotide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a nucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, *Arabidopsis*, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry,

raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In yet another aspect the invention is a transgenic plant comprising one or more of the above described polynucleotides wherein the encoded polypeptide is expressed and regulates transcription of a gene.

In a further aspect the invention provides a method of using the polynucleotide composition to breed a progeny plant from a transgenic plant including crossing plants, producing seeds from transgenic plants, and methods of breeding using transgenic plants, the method comprising transforming a plant with the polynucleotide composition to create a transgenic plant, crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

In a further aspect, the invention provides a progeny plant derived from a parental plant wherein said progeny plant exhibits at least three fold greater messenger RNA levels than said parental plant, wherein the messenger RNA encodes a DNA-binding protein which is capable of binding to a DNA regulatory sequence and inducing expression of a plant trait gene, wherein the progeny plant is characterized by a change in the plant trait compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least ten fold greater messenger RNA levels compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least fifty fold greater messenger RNA levels compared to said parental plant.

In a further aspect, the invention relates to a cloning or expression vector comprising the isolated or recombinant polynucleotide described above or cells comprising the cloning or expression vector.

In yet a further aspect, the invention relates to a composition produced by incubating a polynucleotide of the invention with a nuclease, a restriction enzyme, a polymerase; a polymerase and a primer; a cloning vector, or with a cell.

Furthermore, the invention relates to a method for producing a plant having a modified trait. The method comprises altering the expression of an isolated or recombinant polynucleotide of the invention or altering the expression or activity of a polypeptide of the invention in a plant to produce a modified plant, and selecting the modified plant for a modified trait. In one aspect, the plant is a monocot plant. In another aspect, the plant is a dicot plant. In another aspect the recombinant polynucleotide is from a dicot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a dicot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a dicot plant and the plant is a dicot plant.

In another aspect, the invention is a transgenic plant comprising an isolated or recombinant polynucleotide encoding a polypeptide wherein the polypeptide is selected from the group consisting of SEQ ID NOs: 2 - 2N, where $N = 2-561$. In yet another aspect, the invention is a plant with altered expression levels of a polypeptide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a polynucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, Arabidopsis, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the

invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In another aspect, the invention relates to a method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of the invention. The method comprises expressing a polypeptide encoded by the polynucleotide in a plant; and identifying at least one factor that is modulated by or interacts with the polypeptide. In one embodiment the method for identifying modulating or interacting factors is by detecting binding by the polypeptide to a promoter sequence, or by detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system, or by detecting expression of a factor by hybridization to a microarray, subtractive hybridization, or differential display.

In yet another aspect, the invention is a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest. The method comprises placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of the invention and monitoring one or more of the expression level of the polynucleotide in the plant, the expression level of the polypeptide in the plant, and modulation of an activity of the polypeptide in the plant.

In yet another aspect, the invention relates to an integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of the invention, or to a polypeptide encoded by the polynucleotide. The integrated system, computer or computer readable medium may comprise a link between one or more sequence strings to a modified plant trait.

In yet another aspect, the invention is a method for identifying a sequence similar or homologous to one or more polynucleotides of the invention, or one or more polypeptides encoded by the polynucleotides. The method comprises providing a sequence database, and querying the sequence database with one or more target sequences corresponding to the one or more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

The method may further comprise of linking the one or more of the polynucleotides of the invention, or encoded polypeptides, to a modified plant phenotype.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING, TABLES, AND FIGURE

The Sequence Listing provides exemplary polynucleotide and polypeptide sequences of the invention. The traits associated with the use of the sequences are included in the Examples.

Diskette1 is a read-only memory computer-readable diskette and contains a copy of the Sequence Listing in ASCII text format. The Sequence Listing is named "SEQLIST514442002041" and is 929 kilobytes in size. The copy of the Sequence Listing on the diskette is hereby incorporated by reference in its entirety.

Table 4 shows the polynucleotides and polypeptides identified by SEQ ID NO; Mendel Gene ID No.; conserved domain of the polypeptide; and if the polynucleotide was tested in a transgenic assay. The first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Table 5 lists a summary of orthologous and homologous sequences identified using BLAST (tblastx program). The first column shows the polynucleotide sequence identifier (SEQ ID NO), the second column shows the corresponding cDNA identifier (Gene ID), the third column shows the orthologous or homologous polynucleotide GenBank Accession Number (Test Sequence ID), the fourth column shows the

calculated probability value that the sequence identity is due to chance (Smallest Sum Probability), the fifth column shows the plant species from which the test sequence was isolated (Test Sequence Species), and the sixth column shows the orthologous or homologous test sequence GenBank annotation (Test Sequence GenBank Annotation).

Figure 1 shows a phylogenetic tree of related plant families adapted from Daly et al. (2001 *Plant Physiology* 127:1328-1333).

Detailed Description of Exemplary Embodiments

In an important aspect, the present invention relates to polynucleotides and polypeptides, e.g. for modifying phenotypes of plants. Throughout this disclosure, various information sources are referred to and/or are specifically incorporated. The information sources include scientific journal articles, patent documents, textbooks, and World Wide Web browser-inactive page addresses, for example. While the reference to these information sources clearly indicates that they can be used by one of skill in the art, applicants specifically incorporate each and every one of the information sources cited herein, in their entirety, whether or not a specific mention of "incorporation by reference" is noted. The contents and teachings of each and every one of the information sources can be relied on and used to make and use embodiments of the invention.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a plant" includes a plurality of such plants, and a reference to "a stress" is a reference to one or more stresses and equivalents thereof known to those skilled in the art, and so forth.

The polynucleotide sequences of the invention encode polypeptides that are members of well-known transcription factor families, including plant transcription factor families, as disclosed in Table 4. Generally, the transcription factors encoded by the present sequences are involved in cell differentiation and proliferation and the regulation of growth. Accordingly, one skilled in the art would recognize that by expressing the present sequences in a plant, one may change the expression of

autologous genes or induce the expression of introduced genes. By affecting the expression of similar autologous sequences in a plant that have the biological activity of the present sequences, or by introducing the present sequences into a plant, one may alter a plant's phenotype to one with improved traits. The sequences of the invention may also be used to transform a plant and introduce desirable traits not found in the wild-type cultivar or strain. Plants may then be selected for those that produce the most desirable degree of over- or underexpression of target genes of interest and coincident trait improvement.

The sequences of the present invention may be from any species, particularly plant species, in a naturally occurring form or from any source whether natural, synthetic, semi-synthetic or recombinant. The sequences of the invention may also include fragments of the present amino acid sequences. In this context, a "fragment" refers to a fragment of a polypeptide sequence which is at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity of a transcription factor. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

As one of ordinary skill in the art recognizes, transcription factors can be identified by the presence of a region or domain of structural similarity or identity to a specific consensus sequence or the presence of a specific consensus DNA-binding site or DNA-binding site motif (see, for example, Riechmann et al., (2000) *Science* 290: 2105-2110). The plant transcription factors may belong to one of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *Biol. Chem.* 379:633-646); the MYB transcription factor family (Martin and Paz-Ares, (1997) *Trends Genet.* 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *Biol. Chem.* 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) *Plant Cell* 4:1575-1588); the zinc finger protein (Z) family (Klug and Schwabe (1995) *FASEB J.* 9: 597-604); the homeobox (HB) protein family (Buerklin in *Guidebook to the Homeobox Genes*, Duboule (ed.) (1994) Oxford University Press);

the CAAT-element binding proteins (Forsburg and Guarente (1989) *Genes Dev.* 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *Mol. Gen. Genet.* 1996 250:7-16); the NAM protein family (Souer et al. (1996) *Cell* 85:159-170); the IAA/AUX proteins (Rouse et al. (1998) *Science* 279:1371-1373); the HLH/MYC protein family (Littlewood et al. (1994) *Prot. Profile* 1:639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) *EMBO J.* 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) *FASEB J.* 8:192-200); the Box P-binding protein (the BPF-1) family (da Costa e Silva et al. (1993) *Plant J.* 4:125-135); the high mobility group (HMG) family (Bustin and Reeves (1996) *Prog. Nucl. Acids Res. Mol. Biol.* 54:35-100); the scarecrow (SCR) family (Di Laurenzio et al. (1996) *Cell* 86:423-433); the GF14 family (Wu et al. (1997) *Plant Physiol.* 114:1421-1431); the polycomb (PCOMB) family (Kennison (1995) *Annu. Rev. Genet.* 29:289-303); the teosinte branched (TEO) family (Luo et al. (1996) *Nature* 383:794-799); the ABI3 family (Giraudat et al. (1992) *Plant Cell* 4:1251-1261); the triple helix (TH) family (Dehesh et al. (1990) *Science* 250:1397-1399); the EIL family (Chao et al. (1997) *Cell* 89:1133-44); the AT-HOOK family (Reeves and Nissen (1990) *J. Biol. Chem.* 265:8573-8582); the S1FA family (Zhou et al. (1995) *Nucleic Acids Res.* 23:1165-1169); the bZIPT2 family (Lu and Ferl (1995) *Plant Physiol.* 109:723); the YABBY family (Bowman et al. (1999) *Development* 126:2387-96); the PAZ family (Bohmert et al. (1998) *EMBO J.* 17:170-80); a family of miscellaneous (MISC) transcription factors including the DPBF family (Kim et al. (1997) *Plant J.* 11:1237-1251) and the SPF1 family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the golden (GLD) family (Hall et al. (1998) *Plant Cell* 10:925-936), the TUBBY family (Boggin et al. (1999) *Science* 286:2119-2125), the heat shock family (Wu C (1995) *Annu Rev Cell Dev Biol* 11:441-469), the ENBP family (Christiansen et al (1996) *Plant Mol Biol* 32:809-821), the RING-zinc family (Jensen et al. (1998) *FEBS letters* 436:283-287), the PDBP family (Janik et al *Virology.* (1989) 168:320-329), the PCF family (Cubas P, et al. *Plant J.* (1999) 18:215-22), the SRS (SHI-related) family (Fridborg et al *Plant Cell* (1999) 11:1019-1032), the CPP (cysteine-rich polycomb-like) family (Cvitanich et al *Proc. Natl. Acad. Sci. U S A.* (2000) 97:8163-8168), the ARF (auxin response factor) family (Ulmasov, et al. (1999) *Proc. Natl. Acad. Sci. USA* 96: 5844-5849), the SWI/SNF family (Collingwood et al *J. Mol. End.* 23:255-275), the ACBF family (Seguin et al (1997) *Plant Mol Biol.* 35:281-291), PCGL (CG-1 like) family (da Costa e Silva et al.

(1994) *Plant Mol Biol.* 25:921-924) the ARID family (Vazquez et al. (1999) *Development.* 126: 733-42), the Jumonji family, Balciunas et al (2000, *Trends Biochem Sci.* 25: 274-276), the bZIP-NIN family (Schauser et al (1999) *Nature* 402: 191-195), the E2F family Kaelin et al (1992) *Cell* 70: 351-364) and the GRF-like family (Knaap et al (2000) *Plant Physiol.* 122: 695-704). As indicated by any part of the list above and as known in the art, transcription factors have been sometimes categorized by class, family, and sub-family according to their structural content and consensus DNA-binding site motif, for example. Many of the classes and many of the families and sub-families are listed here. However, the inclusion of one sub-family and not another, or the inclusion of one family and not another, does not mean that the invention does not encompass polynucleotides or polypeptides of a certain family or sub-family. The list provided here is merely an example of the types of transcription factors and the knowledge available concerning the consensus sequences and consensus DNA-binding site motifs that help define them as known to those of skill in the art (each of the references noted above are specifically incorporated herein by reference). A transcription factor may include, but is not limited to, any polypeptide that can activate or repress transcription of a single gene or a number of genes. This polypeptide group includes, but is not limited to, DNA-binding proteins, DNA-binding protein binding proteins, protein kinases, protein phosphatases, GTP-binding proteins, and receptors, and the like.

In addition to methods for modifying a plant phenotype by employing one or more polynucleotides and polypeptides of the invention described herein, the polynucleotides and polypeptides of the invention have a variety of additional uses. These uses include their use in the recombinant production (i.e., expression) of proteins; as regulators of plant gene expression, as diagnostic probes for the presence of complementary or partially complementary nucleic acids (including for detection of natural coding nucleic acids); as substrates for further reactions, e.g., mutation reactions, PCR reactions, or the like; as substrates for cloning e.g., including digestion or ligation reactions; and for identifying exogenous or endogenous modulators of the transcription factors. A "polynucleotide" is a nucleic acid sequence comprising a plurality of polymerized nucleotides, e.g., at least about 15 consecutive polymerized nucleotides, optionally at least about 30 consecutive nucleotides, at least about 50 consecutive nucleotides. In many instances, a polynucleotide comprises a nucleotide

sequence encoding a polypeptide (or protein) or a domain or fragment thereof. Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker, or the like. The polynucleotide can be single stranded or double stranded DNA or RNA. The polynucleotide optionally comprises modified bases or a modified backbone. The polynucleotide can be, e.g., genomic DNA or RNA, a transcript (such as an mRNA), a cDNA, a PCR product, a cloned DNA, a synthetic DNA or RNA, or the like. The polynucleotide can comprise a sequence in either sense or antisense orientations.

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide comprises a nucleotide sequence not found in nature, or the polynucleotide is in a context other than that in which it is naturally found, e.g., separated from nucleotide sequences with which it typically is in proximity in nature, or adjacent (or contiguous with) nucleotide sequences with which it typically is not in proximity. For example, the sequence at issue can be cloned into a vector, or otherwise recombined with one or more additional nucleic acid.

An "isolated polynucleotide" is a polynucleotide whether naturally occurring or recombinant, that is present outside the cell in which it is typically found in nature, whether purified or not. Optionally, an isolated polynucleotide is subject to one or more enrichment or purification procedures, e.g., cell lysis, extraction, centrifugation, precipitation, or the like.

A "polypeptide" is an amino acid sequence comprising a plurality of consecutive polymerized amino acid residues e.g., at least about 15 consecutive polymerized amino acid residues, optionally at least about 30 consecutive polymerized amino acid residues, at least about 50 consecutive polymerized amino acid residues. In many instances, a polypeptide comprises a polymerized amino acid residue sequence that is a transcription factor or a domain or portion or fragment thereof. Additionally, the polypeptide may comprise a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA-binding domain, or the like. The polypeptide optionally comprises modified

amino acid residues, naturally occurring amino acid residues not encoded by a codon, non-naturally occurring amino acid residues.

A "recombinant polypeptide" is a polypeptide produced by translation of a recombinant polynucleotide. A "synthetic polypeptide" is a polypeptide created by consecutive polymerization of isolated amino acid residues using methods well known in the art. An "isolated polypeptide," whether a naturally occurring or a recombinant polypeptide, is more enriched in (or out of) a cell than the polypeptide in its natural state in a wild type cell, e.g., more than about 5% enriched, more than about 10% enriched, or more than about 20%, or more than about 50%, or more, enriched, i.e., alternatively denoted: 105%, 110%, 120%, 150% or more, enriched relative to wild type standardized at 100%. Such an enrichment is not the result of a natural response of a wild type plant. Alternatively, or additionally, the isolated polypeptide is separated from other cellular components with which it is typically associated, e.g., by any of the various protein purification methods herein.

"Identity" or "similarity" refers to sequence similarity between two polynucleotide sequences or between two polypeptide sequences, with identity being a more strict comparison. The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more polynucleotide sequences or two or more polypeptide sequences. Identity or similarity can be determined by comparing a position in each sequence that may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base or amino acid, then the molecules are identical at that position. A degree of similarity or identity between polynucleotide sequences is a function of the number of identical or matching nucleotides at positions shared by the polynucleotide sequences. A degree of identity of polypeptide sequences is a function of the number of identical amino acids at positions shared by the polypeptide sequences. A degree of homology or similarity of polypeptide sequences is a function of the number of amino acids, i.e., structurally related, at positions shared by the polypeptide sequences.

"Altered" nucleic acid sequences encoding polypeptide include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting

in a polynucleotide encoding a polypeptide with at least one functional characteristic of the polypeptide. Included within this definition are polymorphisms that may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding polypeptide, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding polypeptide. The encoded polypeptide protein may also be "altered", and may contain deletions, insertions, or substitutions of amino acid residues that produce a silent change and result in a functionally equivalent polypeptide. Deliberate amino acid substitutions may be made on the basis of similarity in residue side chain chemistry, including, but not limited to, polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological activity of polypeptide is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine. Alignments between different polypeptide sequences may be used to calculate "percentage sequence similarity".

The term "plant" includes whole plants, shoot vegetative organs/structures (e.g., leaves, stems and tubers), roots, flowers and floral organs/structures (e.g., bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (e.g., vascular tissue, ground tissue, and the like) and cells (e.g., guard cells, egg cells, and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, horsetails, psilophytes, lycophytes, bryophytes, and multicellular algae. (See for example, Figure 1, adapted from Daly et al. 2001 *Plant Physiology* 127:1328-1333; and see also Tudge, C., The Variety of Life, Oxford University Press, New York, 2000, pp. 547-606.)

A "transgenic plant" refers to a plant that contains genetic material not found in a wild type plant of the same species, variety or cultivar. The genetic material may

include a transgene, an insertional mutagenesis event (such as by transposon or T-DNA insertional mutagenesis), an activation tagging sequence, a mutated sequence, a homologous recombination event or a sequence modified by chimeraplasty. Typically, the foreign genetic material has been introduced into the plant by human manipulation, but any method can be used as one of skill in the art recognizes.

A transgenic plant may contain an expression vector or cassette. The expression cassette typically comprises a polypeptide-encoding sequence operably linked (i.e., under regulatory control of) to appropriate inducible or constitutive regulatory sequences that allow for the expression of polypeptide. The expression cassette can be introduced into a plant by transformation or by breeding after transformation of a parent plant. A plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, e.g., a plant explant, as well as to progeny thereof, and to *in vitro* systems that mimic biochemical or cellular components or processes in a cell.

"Ectopic expression or altered expression" in reference to a polynucleotide indicates that the pattern of expression in, e.g., a transgenic plant or plant tissue, is different from the expression pattern in a wild type plant or a reference plant of the same species. The pattern of expression may also be compared with a reference expression pattern in a wild type plant of the same species. For example, the polynucleotide or polypeptide is expressed in a cell or tissue type other than a cell or tissue type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to altered expression patterns that are produced by lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern can be transient or stable, constitutive or inducible. In reference to a polypeptide, the term "ectopic expression or altered expression" further may relate to altered activity levels resulting from the interactions of the polypeptides with exogenous or endogenous modulators or from interactions with factors or as a result of the chemical modification of the polypeptides.

A "fragment" or "domain," with respect to a polypeptide, refers to a subsequence of the polypeptide. In some cases, the fragment or domain, is a subsequence of the polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA-binding site or domain that binds to a DNA promoter region, an activation domain, or a domain for protein-protein interactions. Fragments can vary in size from as few as 6 amino acids to the full length of the intact polypeptide, but are preferably at least about 30 amino acids in length and more preferably at least about 60 amino acids in length. In reference to a polynucleotide sequence, "a fragment" refers to any subsequence of a polynucleotide, typically, of at least about 15 consecutive nucleotides, preferably at least about 30 nucleotides, more preferably at least about 50 nucleotides, of any of the sequences provided herein.

The invention also encompasses production of DNA sequences that encode transcription factors and transcription factor derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding transcription factors or any fragment thereof.

A "conserved domain", with respect to a polypeptide, refers to a domain within a transcription factor family which exhibits a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% amino acid residue sequence identity of a polypeptide of consecutive amino acid residues. A fragment or domain can be referred to as outside a consensus sequence or outside a consensus DNA-binding site that is known to exist or that exists for a particular transcription factor class, family, or sub-family. In this case, the fragment or domain will not include the exact amino acids of a consensus sequence or consensus DNA-binding

site of a transcription factor class, family or sub-family, or the exact amino acids of a particular transcription factor consensus sequence or consensus DNA-binding site. Furthermore, a particular fragment, region, or domain of a polypeptide, or a polynucleotide encoding a polypeptide, can be "outside a conserved domain" if all the amino acids of the fragment, region, or domain fall outside of a defined conserved domain(s) for a polypeptide or protein. The conserved domains for each of polypeptides of SEQ ID NOs:2 - 2N, where N = 2-561, are listed in Table 4 as described in Example VII. Also, many of the polypeptides of Table 4 have conserved domains specifically indicated by start and stop sites. A comparison of the regions of the polypeptides in SEQ ID NOs:2 - 2N, where N = 2-561, or of those in Table 4, allows one of skill in the art to identify conserved domain(s) for any of the polypeptides listed or referred to in this disclosure, including those in Table 4.

A "trait" refers to a physiological, morphological, biochemical, or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic is visible to the human eye, such as seed or plant size, or can be measured by biochemical techniques, such as detecting the protein, starch, or oil content of seed or leaves, or by observation of a metabolic or physiological process, e.g. by measuring uptake of carbon dioxide, or by the observation of the expression level of a gene or genes, e.g., by employing Northern analysis, RT-PCR, microarray gene expression assays, or reporter gene expression systems, or by agricultural observations such as stress tolerance, yield, or pathogen tolerance. Any technique can be used to measure the amount of, comparative level of, or difference in any selected chemical compound or macromolecule in the transgenic plants, however.

"Trait modification" refers to a detectable difference in a characteristic in a plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. In some cases, the trait modification can be evaluated quantitatively. For example, the trait modification can entail at least about a 2% increase or decrease in an observed trait (difference), at least a 5% difference, at least about a 10% difference, at least about a 20% difference, at least about a 30%, at least about a 50%, at least about a 70%, or at least about a 100%, or an even greater difference compared with a wild type plant. It is known that there can be a natural variation in the modified trait. Therefore, the trait modification

observed entails a change of the normal distribution of the trait in the plants compared with the distribution observed in wild type plant.

I. Traits Which May Be Modified

Trait modifications of particular interest include those to seed (such as embryo or endosperm), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; improved tolerance to microbial, fungal or viral diseases; improved tolerance to pest infestations, including nematodes, mollicutes, parasitic higher plants or the like; decreased herbicide sensitivity; improved tolerance of heavy metals or enhanced ability to take up heavy metals; improved growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that can be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenillipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that can be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves, inflorescences, and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that can be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

II. Transcription Factors Modify Expression Of Endogenous Genes

Expression of genes which encode transcription factors that modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription

factors may also modify expression of endogenous genes, polynucleotides, and proteins. Examples include Peng et al. (1997, *Genes and Development* 11:3194-3205) and Peng et al. (1999, *Nature*, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, *Plant Cell* 13:1791-1802); Nandi et al. (2000, *Curr. Biol.* 10:215-218); Coupland (1995, *Nature* 377:482-483); and Weigel and Nilsson (1995, *Nature* 377:482-500).

In another example, Mandel et al. (1992, *Cell* 71:133-143) and Suzuki et al. (2001, *Plant J.* 28:409-418) teach that a transcription factor expressed in another plant species elicits the same or very similar phenotypic response of the endogenous sequence, as often predicted in earlier studies of *Arabidopsis* transcription factors in *Arabidopsis* (see Mandel et al., 1992, *supra*; Suzuki et al., 2001, *supra*).

Other examples include Müller et al. (2001, *Plant J.* 28:169-179); Kim et al. (2001, *Plant J.* 25:247-259); Kyoizuka and Shimamoto (2002, *Plant Cell Physiol.* 43:130-135); Boss and Thomas (2002, *Nature*, 416:847-850); He et al. (2000, *Transgenic Res.*, 9:223-227); and Robson et al. (2001, *Plant J.* 28:619-631).

In yet another example, Gilmour et al. (1998, *Plant J.* 16:433-442) teach an *Arabidopsis* AP2 transcription factor, CBF1, which, when overexpressed in transgenic plants, increases plant freezing tolerance. Jaglo et al (2001, *Plant Physiol.* 127:910-917) further identified sequences in *Brassica napus* which encode CBF-like genes and that transcripts for these genes accumulated rapidly in response to low temperature. Transcripts encoding CBF-like proteins were also found to accumulate rapidly in response to low temperature in wheat, as well as in tomato. An alignment of the CBF proteins from *Arabidopsis*, *B. napus*, wheat, rye, and tomato revealed the presence of conserved amino acid sequences, PKK/RPAGR_xKFxETRHP and DSAWR, that bracket the AP2/EREBP DNA binding domains of the proteins and distinguish them from other members of the AP2/EREBP protein family. (See Jaglo et al., *supra*.)

III. Polypeptides and Polynucleotides of the Invention

The present invention provides, among other things, transcription factors (TFs), and transcription factor homologue polypeptides, and isolated or recombinant polynucleotides encoding the polypeptides, or novel variant polypeptides or polynucleotides encoding novel variants of transcription factors derived from the specific sequences provided here. These polypeptides and polynucleotides may be employed to modify a plant's characteristic.

Exemplary polynucleotides encoding the polypeptides of the invention were identified in the *Arabidopsis thaliana* GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. In addition, further exemplary polynucleotides encoding the polypeptides of the invention were identified in the plant GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. Polynucleotide sequences meeting such criteria were confirmed as transcription factors.

Additional polynucleotides of the invention were identified by screening *Arabidopsis thaliana* and/or other plant cDNA libraries with probes corresponding to known transcription factors under low stringency hybridization conditions. Additional sequences, including full length coding sequences were subsequently recovered by the rapid amplification of cDNA ends (RACE) procedure, using a commercially available kit according to the manufacturer's instructions. Where necessary, multiple rounds of RACE are performed to isolate 5' and 3' ends. The full length cDNA was then recovered by a routine end-to-end polymerase chain reaction (PCR) using primers specific to the isolated 5' and 3' ends. Exemplary sequences are provided in the Sequence Listing.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor or knockout plants and the changes in the characteristic(s) or trait(s) of

the plants observed. Therefore, the polynucleotides and polypeptides can be employed to improve the characteristics of plants.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor plant cells and the changes in the expression levels of a number of genes, polynucleotides, and/or proteins of the plant cells observed. Therefore, the polynucleotides and polypeptides can be employed to change expression levels of a genes, polynucleotides, and/or proteins of plants.

IV. Producing Polypeptides

The polynucleotides of the invention include sequences that encode transcription factors and transcription factor homologue polypeptides and sequences complementary thereto, as well as unique fragments of coding sequence, or sequence complementary thereto. Such polynucleotides can be, e.g., DNA or RNA, e.g., mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, oligonucleotides, etc. The polynucleotides are either double-stranded or single-stranded, and include either, or both sense (i.e., coding) sequences and antisense (i.e., non-coding, complementary) sequences. The polynucleotides include the coding sequence of a transcription factor, or transcription factor homologue polypeptide, in isolation, in combination with additional coding sequences (e.g., a purification tag, a localization signal, as a fusion-protein, as a pre-protein, or the like), in combination with non-coding sequences (e.g., introns or inteins, regulatory elements such as promoters, enhancers, terminators, and the like), and/or in a vector or host environment in which the polynucleotide encoding a transcription factor or transcription factor homologue polypeptide is an endogenous or exogenous gene.

A variety of methods exist for producing the polynucleotides of the invention. Procedures for identifying and isolating DNA clones are well known to those of skill in the art, and are described in, e.g., Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds.,

Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

Alternatively, polynucleotides of the invention, can be produced by a variety of in vitro amplification methods adapted to the present invention by appropriate selection of specific or degenerate primers. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qbeta-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger (*supra*), Sambrook (*supra*), and Ausubel (*supra*), as well as Mullis et al., (1987) PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. *See, e.g., Ausubel, Sambrook and Berger, all supra.*

Alternatively, polynucleotides and oligonucleotides of the invention can be assembled from fragments produced by solid-phase synthesis methods. Typically, fragments of up to approximately 100 bases are individually synthesized and then enzymatically or chemically ligated to produce a desired sequence, e.g., a polynucleotide encoding all or part of a transcription factor. For example, chemical synthesis using the phosphoramidite method is described, e.g., by Beaucage et al. (1981) Tetrahedron Letters 22:1859-1869; and Matthes et al. (1984) EMBO J. 3:801-805. According to such methods, oligonucleotides are synthesized, purified, annealed to their complementary strand, ligated and then optionally cloned into suitable vectors. And if so desired, the polynucleotides and polypeptides of the invention can be custom ordered from any of a number of commercial suppliers.

V. Homologous Sequences

Sequences homologous, i.e., that share significant sequence identity or similarity, to those provided in the Sequence Listing, derived from *Arabidopsis thaliana* or from other plants of choice are also an aspect of the invention.

Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, rape, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, Brussels sprouts, and kohlrabi). Other crops, fruits and vegetables whose phenotype can be changed include barley, rye, millet, sorghum, currant, avocado, citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato, and beans. The homologous sequences may also be derived from woody species, such pine, poplar and eucalyptus, or mint or other labiates.

Orthologs And Paralogs

Several different methods are known by those of skill in the art for identifying and defining these functionally homologous sequences. Three general methods for defining paralog and ortholog are described; a paralog or ortholog or homolog may be identified by one or more of the methods described below.

Orthologs and paralog are evolutionarily related genes that have similar sequence and similar functions. Orthologs are structurally related genes in different species that are derived from a speciation event. Paralog are structurally related genes within a single species that are derived by a duplication event.

Within a single plant species, gene duplication may cause two copies of a particular gene, giving rise to two or more genes with similar sequence and similar function known as paralog. A paralog is therefore a similar gene with a similar function within the same species. Paralog typically cluster together or in the same

clade (a group of similar genes) when a gene family phylogeny is analyzed using programs such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266 383-402). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle (1987) *J. Mol. Evol.* 25:351-360). For example, a clade of very similar MADS domain transcription factors from *Arabidopsis* all share a common function in flowering time (Ratcliffe et al. (2001) *Plant Physiol.* 126:122-132), and a group of very similar AP2 domain transcription factors from *Arabidopsis* are involved in tolerance of plants to freezing (Gilmour et al. (1998) *Plant J.* 16:433-442). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous or orthologous sequences that share the same function. (See also, for example, Mount, D.W. (2001) Bioinformatics: Sequence and Genome Analysis Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York page 543.)

Speciation, the production of new species from a parental species, can also give rise to two or more genes with similar sequence and similar function. These genes, termed orthologs, often have an identical function within their host plants and are often interchangeable between species without losing function. Because plants have common ancestors, many genes in any plant species will have a corresponding orthologous gene in another plant species. Once a phylogenetic tree for a gene family of one species has been constructed using a program such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266:383-402), potential orthologous sequences can be placed into the phylogenetic tree and its relationship to genes from the species of interest can be determined. Once the ortholog pair has been identified, the function of the test ortholog can be determined by determining the function of the reference ortholog.

Transcription factors that are homologous to the listed sequences will typically share at least about 30% amino acid sequence identity, or at least about 30% amino acid sequence identity outside of a known consensus sequence or consensus DNA-binding site. More closely related transcription factors can share at least about 50%,

about 60%, about 65%, about 70%, about 75% or about 80% or about 90% or about 95% or about 98% or more sequence identity with the listed sequences, or with the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site, or with the listed sequences excluding one or all conserved domain. Factors that are most closely related to the listed sequences share, e.g., at least about 85%, about 90% or about 95% or more % sequence identity to the listed sequences, or to the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site or outside one or all conserved domain. At the nucleotide level, the sequences will typically share at least about 40% nucleotide sequence identity, preferably at least about 50%, about 60%, about 70% or about 80% sequence identity, and more preferably about 85%, about 90%, about 95% or about 97% or more sequence identity to one or more of the listed sequences, or to a listed sequence but excluding or outside a known consensus sequence or consensus DNA-binding site, or outside one or all conserved domain. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein. Conserved domains within a transcription factor family may exhibit a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity. Transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence identity over the entire length of the polypeptide or the homolog. In addition, transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence similarity over the entire length of the polypeptide or the homolog.

Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Inc. Madison, Wis.). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D. G. and P. M. Sharp (1988) Gene

73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. Other alignment algorithms or programs may be used, including FASTA, BLAST, or ENTREZ, FASTA and BLAST. These are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with or without default settings. ENTREZ is available through the National Center for Biotechnology Information. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as if it were a single amino acid or nucleotide mismatch between the two sequences (see USPN 6,262,333).

Other techniques for alignment are described in *Methods in Enzymology*, vol. 266: *Computer Methods for Macromolecular Sequence Analysis* (1996), ed. Doolittle, Academic Press, Inc., San Diego, Calif., USA. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Methods Mol. Biol.* 70: 173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to pick up distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Nucleic acid-encoded amino acid sequences can be used to search both protein and DNA databases.

The percentage similarity between two polypeptide sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between polynucleotide sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods

known in the art, e.g., by varying hybridization conditions (see US Patent Application No. 20010010913).

Thus, the invention provides methods for identifying a sequence similar or paralogous or orthologous or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

In addition, one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to search against a BLOCKS (Bairoch et al. (1997) *Nucleic Acids Res.* 25:217-221), PFAM, and other databases which contain previously identified and annotated motifs, sequences and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith et al. (1992) *Protein Engineering* 5:35-51) as well as algorithms such as Basic Local Alignment Search Tool (BLAST; Altschul, S. F. (1993) *J. Mol. Evol.* 36:290-300; Altschul et al. (1990) *supra*), BLOCKS (Henikoff, S. and Henikoff, G. J. (1991) *Nucleic Acids Research* 19:6565-6572), Hidden Markov Models (HMM; Eddy, S. R. (1996) *Cur. Opin. Str. Biol.* 6:361-365; Sonnhammer et al. (1997) *Proteins* 28:405-420), and the like, can be used to manipulate and analyze polynucleotide and polypeptide sequences encoded by polynucleotides. These databases, algorithms and other methods are well known in the art and are described in Ausubel et al. (1997; *Short Protocols in Molecular Biology*, John Wiley & Sons, New York N.Y., unit 7.7) and in Meyers, R. A. (1995; *Molecular Biology and Biotechnology*, Wiley VCH, New York N.Y., p 856-853).

Furthermore, methods using manual alignment of sequences similar or homologous to one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to identify regions of similarity and conserved domains. Such manual methods are well-known of those of skill in the art and can include, for example, comparisons of tertiary structure between a

polypeptide sequence encoded by a polynucleotide which comprises a known function with a polypeptide sequence encoded by a polynucleotide sequence which has a function not yet determined. Such examples of tertiary structure may comprise predicted alpha helices, beta-sheets, amphipathic helices, leucine zipper motifs, zinc finger motifs, proline-rich regions, cysteine repeat motifs, and the like.

VI. Identifying Polynucleotides or Nucleic Acids by Hybridization

Polynucleotides homologous to the sequences illustrated in the Sequence Listing and tables can be identified, e.g., by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in the references cited above. Encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NOs: 860; 802; 240; 274; 558; 24; 1120; 44; 460; 286; 120; 130; 134; 698; 832; 580; 612; 48, and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507-511.) Estimates of homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions.

In addition to the nucleotide sequences listed in Tables 4 and 5, full length cDNA, orthologs, paralogs and homologs of the present nucleotide sequences may be

identified and isolated using well known methods. The cDNA libraries orthologs, paralogs and homologs of the present nucleotide sequences may be screened using hybridization methods to determine their utility as hybridization target or amplification probes.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions, e.g., to a unique subsequence, of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C. For example, high stringency is about 0.2 x SSC, 0.1% SDS at 65° C. Ultra-high stringency will be the same conditions except the wash temperature is raised about 3 to about 5° C, and ultra-ultra-high stringency will be the same conditions except the wash temperature is raised about 6 to about 9° C. For identification of less closely related homologues washes can be performed at a lower temperature, e.g., 50° C. In general, stringency is increased by raising the wash temperature and/or decreasing the concentration of SSC, as known in the art.

In another example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the

art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps that follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25° C, more preferably of at least about 42° C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65° C. The most preferred high stringency washes are of at least about 68° C. For example, in a preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, the wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art (see U.S. Patent Application No. 20010010913).

As another example, stringent conditions can be selected such that an oligonucleotide that is perfectly complementary to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 5-10x higher signal to noise ratio than the ratio for hybridization of the perfectly complementary oligonucleotide to a nucleic acid encoding a transcription factor known as of the filing

date of the application. Conditions can be selected such that a higher signal to noise ratio is observed in the particular assay which is used, e.g., about 15x, 25x, 35x, 50x or more. Accordingly, the subject nucleic acid hybridizes to the unique coding oligonucleotide with at least a 2x higher signal to noise ratio as compared to hybridization of the coding oligonucleotide to a nucleic acid encoding known polypeptide. Again, higher signal to noise ratios can be selected, e.g., about 5x, 10x, 25x, 35x, 50x or more. The particular signal will depend on the label used in the relevant assay, e.g., a fluorescent label, a colorimetric label, a radioactive label, or the like.

Alternatively, transcription factor homolog polypeptides can be obtained by screening an expression library using antibodies specific for one or more transcription factors. With the provision herein of the disclosed transcription factor, and transcription factor homologue nucleic acid sequences, the encoded polypeptide(s) can be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the polypeptide(s) in question. Antibodies can also be raised against synthetic peptides derived from transcription factor, or transcription factor homologue, amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone additional transcription factor homologues, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzymatic activity.

VII. Sequence Variations

It will readily be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the transcription factors and transcription factor homologue polypeptides of the invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing. Nucleic acids having a sequence that differs from the sequences shown in the Sequence Listing, or complementary sequences, that encode functionally equivalent peptides (i.e., peptides having some degree of equivalent or similar

biological activity) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code, are also within the scope of the invention.

Altered polynucleotide sequences encoding polypeptides include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide encoding a polypeptide with at least one functional characteristic of the instant polypeptides. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding the instant polypeptides, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding the instant polypeptides.

Allelic variant refers to any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (i.e., no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene. Splice variant refers to alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Those skilled in the art would recognize that the polypeptide sequence G681, SEQ ID NO: 580, represents a single transcription factor; allelic variation and alternative splicing may be expected to occur. Allelic variants of the polypeptide sequence of SEQ ID NO: 579 can be cloned by probing cDNA or genomic libraries from different individual organisms according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO: 579, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants

of SEQ ID NO: 580. cDNAs generated from alternatively spliced mRNAs, which retain the properties of the transcription factor are included within the scope of the present invention, as are polypeptides encoded by such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individual organisms or tissues according to standard procedures known in the art (see USPN 6,388,064).

For example, Table 1 illustrates, e.g., that the codons AGC, AGT, TCA, TCC, TCG, and TCT all encode the same amino acid: serine. Accordingly, at each position in the sequence where there is a codon encoding serine, any of the above trinucleotide sequences can be used without altering the encoded polypeptide.

Table 1

Amino acid			Possible Codons							
Alanine	Ala	A	GCA	GCC	GCG	GCU				
Cysteine	Cys	C	TGC	TGT						
Aspartic acid	Asp	D	GAC	GAT						
Glutamic acid	Glu	E	GAA	GAG						
Phenylalanine	Phe	F	TTC	TTT						
Glycine	Gly	G	GGA	GGC	GGG	GGT				
Histidine	His	H	CAC	CAT						
Isoleucine	Ile	I	ATA	ATC	ATT					
Lysine	Lys	K	AAA	AAG						
Leucine	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT		
Methionine	Met	M	ATG							
Asparagine	Asn	N	AAC	AAT						
Proline	Pro	P	CCA	CCC	CCG	CCT				
Glutamine	Gln	Q	CAA	CAG						
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGT		
Serine	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT		
Threonine	Thr	T	ACA	ACC	ACG	ACT				
Valine	Val	V	GTA	GTC	GTG	GTT				
Tryptophan	Trp	W	TGG							
Tyrosine	Tyr	Y	TAC	TAT						

Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed "silent" variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence selected from the above table are a feature of the invention.

In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide, these conservative variants are, likewise, a feature of the invention.

For example, substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) Meth. Enzymol. (1993) vol. 217, Academic Press) or the other methods noted below. Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function.

Conservative substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 2 when it is desired to maintain the activity of the protein. Table 2 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

Table 2

Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Gln	Asn
Cys	Ser
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr; Gly
Thr	Ser; Val
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Similar substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 3 when it is desired to maintain the activity of the protein. Table 3 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as structural and functional substitutions. For example, a residue in column 1 of Table 3

may be substituted with residue in column 2; in addition, a residue in column 2 of Table 3 may be substituted with the residue of column 1.

Table 3

Residue	Similar Substitutions
Ala	Ser; Thr; Gly; Val; Leu; Ile
Arg	Lys; His; Gly
Asn	Gln; His; Gly; Ser; Thr
Asp	Glu; Ser; Thr
Gln	Asn; Ala
Cys	Ser; Gly
Glu	Asp
Gly	Pro; Arg
His	Asn; Gln; Tyr; Phe; Lys; Arg
Ile	Ala; Leu; Val; Gly; Met
Leu	Ala; Ile; Val; Gly; Met
Lys	Arg; His; Gln; Gly; Pro
Met	Leu; Ile; Phe
Phe	Met; Leu; Tyr; Trp; His; Val; Ala
Ser	Thr; Gly; Asp; Ala; Val; Ile; His
Thr	Ser; Val; Ala; Gly
Trp	Tyr; Phe; His
Tyr	Trp; Phe; His
Val	Ala; Ile; Leu; Gly; Thr; Ser; Glu

Substitutions that are less conservative than those in Table 2 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are

expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

VIII. Further Modifying Sequences of the Invention – Mutation/Forced Evolution

In addition to generating silent or conservative substitutions as noted, above, the present invention optionally includes methods of modifying the sequences of the Sequence Listing. In the methods, nucleic acid or protein modification methods are used to alter the given sequences to produce new sequences and/or to chemically or enzymatically modify given sequences to change the properties of the nucleic acids or proteins.

Thus, in one embodiment, given nucleic acid sequences are modified, e.g., according to standard mutagenesis or artificial evolution methods to produce modified sequences. The modified sequences may be created using purified natural polynucleotides isolated from any organism or may be synthesized from purified compositions and chemicals using chemical means well known to those of skill in the art. For example, Ausubel, *supra*, provides additional details on mutagenesis methods. Artificial forced evolution methods are described, for example, by Stemmer (1994) Nature 370:389-391, Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751, and U.S. Patents 5,811,238, 5,837,500, and 6,242,568. Methods for engineering synthetic transcription factors and other polypeptides are described, for example, by Zhang et al. (2000) J. Biol. Chem. 275:33850-33860, Liu et al. (2001) J. Biol. Chem. 276:11323-11334, and Isalan et al. (2001) Nature Biotechnol. 19:656-660. Many other mutation and evolution methods are also available and expected to be within the skill of the practitioner.

Similarly, chemical or enzymatic alteration of expressed nucleic acids and polypeptides can be performed by standard methods. For example, sequence can be modified by addition of lipids, sugars, peptides, organic or inorganic compounds, by the inclusion of modified nucleotides or amino acids, or the like. For example, protein modification techniques are illustrated in Ausubel, *supra*. Further details on chemical and enzymatic modifications can be found herein. These modification methods can be used to modify any given sequence, or to modify any sequence produced by the various mutation and artificial evolution modification methods noted herein.

Accordingly, the invention provides for modification of any given nucleic acid by mutation, evolution, chemical or enzymatic modification, or other available methods, as well as for the products produced by practicing such methods, e.g., using the sequences herein as a starting substrate for the various modification approaches.

For example, optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host can be used e.g., to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced using a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *Saccharomyces cerevisiae* and mammals are TAA and TGA, respectively. The preferred stop codon for monocotyledonous plants is TGA, whereas insects and *E. coli* prefer to use TAA as the stop codon.

The polynucleotide sequences of the present invention can also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which modify the sequence to facilitate cloning, processing and/or expression of the gene product. For example, alterations are optionally introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to introduce splice sites, etc.

Furthermore, a fragment or domain derived from any of the polypeptides of the invention can be combined with domains derived from other transcription factors or synthetic domains to modify the biological activity of a transcription factor. For instance, a DNA-binding domain derived from a transcription factor of the invention can be combined with the activation domain of another transcription factor or with a synthetic activation domain. A transcription activation domain assists in initiating transcription from a DNA-binding site. Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; and Aoyama et al. (1995) Plant Cell 7:1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) Cell 51: 113-119) and synthetic peptides (Giniger and Ptashne, (1987) Nature 330:670-672).

IX. Expression and Modification of Polypeptides

Typically, polynucleotide sequences of the invention are incorporated into recombinant DNA (or RNA) molecules that direct expression of polypeptides of the invention in appropriate host cells, transgenic plants, in vitro translation systems, or the like. Due to the inherent degeneracy of the genetic code, nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can be substituted for any listed sequence to provide for cloning and expressing the relevant homologue.

X. Vectors, Promoters, and Expression Systems

The present invention includes recombinant constructs comprising one or more of the nucleic acid sequences herein. The constructs typically comprise a vector, such as a plasmid, a cosmid, a phage, a virus (e.g., a plant virus), a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts that describe molecular biological techniques useful herein, including the use and production of vectors, promoters and many other relevant

topics, include Berger, Sambrook and Ausubel, *supra*. Any of the identified sequences can be incorporated into a cassette or vector, e.g., for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology, Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technology 3: 637-642, for dicotyledonous plants.

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods can involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9: 957-962) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technology 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48, and for *Agrobacterium*-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which can be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which

confers constitutive, high-level expression in most plant tissues (*see, e.g.,* Odell et al. (1985) Nature 313:810-812); the nopaline synthase promoter (An et al. (1988) Plant Physiol 88:547-552); and the octopine synthase promoter (Fromm et al. (1989) Plant Cell 1: 977-984).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of a TF sequence in plants. Choice of a promoter is based largely on the phenotype of interest and is determined by such factors as tissue (e.g., seed, fruit, root, pollen, vascular tissue, flower, carpel, etc.), inducibility (e.g., in response to wounding, heat, cold, drought, light, pathogens, etc.), timing, developmental stage, and the like. Numerous known promoters have been characterized and can favorably be employed to promote expression of a polynucleotide of the invention in a transgenic plant or cell of interest. For example, tissue specific promoters include: seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the *dru 1* promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) Plant Mol Biol 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) Plant Mol Biol 37:977-988), flower-specific (Kaiser et al. (1995) Plant Mol Biol 28:231-243), pollen (Baerson et al. (1994) Plant Mol Biol 26:1947-1959), carpels (Ohl et al. (1990) Plant Cell 2:837-848), pollen and ovules (Baerson et al. (1993) Plant Mol Biol 22:255-267), auxin-inducible promoters (such as that described in van der Kop et al. (1999) Plant Mol Biol 39:979-990 or Baumann et al. (1999) Plant Cell 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) Plant Mol Biol 38:743-753), promoters responsive to gibberellin (Shi et al. (1998) Plant Mol Biol 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley et al. (1993) Plant Mol Biol 22: 13-23), light (e.g., the pea *rbcS-3A* promoter, Kuhlemeier et al. (1989) Plant Cell 1:471, and the maize *rbcS* promoter, Schaffner and Sheen (1991) Plant Cell 3: 997); wounding (e.g., *wun1*, Siebertz et al. (1989) Plant Cell 1: 961); pathogens (such as the PR-1

promoter described in Buchel et al. (1999) Plant Mol. Biol. 40:387-396, and the PDF1.2 promoter described in Manners et al. (1998) Plant Mol. Biol. 38:1071-80), and chemicals such as methyl jasmonate or salicylic acid (Gatz et al. (1997) Plant Mol Biol 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (An and Amazon (1995) Science 270: 1986-1988); or late seed development (Odell et al. (1994) Plant Physiol 106:447-458).

Plant expression vectors can also include RNA processing signals that can be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors can include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Additional Expression Elements

Specific initiation signals can aid in efficient translation of coding sequences. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon can be separately provided. The initiation codon is provided in the correct reading frame to facilitate transcription. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use.

Expression Hosts

The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention (including fragments thereof) by recombinant techniques. Host cells are genetically engineered (i.e., nucleic acids are introduced, e.g., transduced, transformed or transfected) with the vectors of this invention, which may be, for example, a cloning vector or an expression vector comprising the relevant nucleic acids herein. The

vector is optionally a plasmid, a viral particle, a phage, a naked nucleic acid, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the relevant gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, Sambrook and Ausubel.

The host cell can be a eukaryotic cell, such as a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Plant protoplasts are also suitable for some applications. For example, the DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (Fromm et al., (1985) Proc. Natl. Acad. Sci. USA 82, 5824, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., (1982) Molecular Biology of Plant Tumors, (Academic Press, New York) pp. 549-560; US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327, 70-73), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* or *A. rhizogenes* carrying a T-DNA plasmid in which DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and a portion is stably integrated into the plant genome (Horsch et al. (1984) Science 233:496-498; Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80, 4803).

The cell can include a nucleic acid of the invention which encodes a polypeptide, wherein the cells expresses a polypeptide of the invention. The cell can also include vector sequences, or the like. Furthermore, cells and transgenic plants that include any polypeptide or nucleic acid above or throughout this specification, e.g., produced by transduction of a vector of the invention, are an additional feature of the invention.

For long-term, high-yield production of recombinant proteins, stable expression can be used. Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the

expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding mature proteins of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

XI. Modified Amino Acid Residues

Polypeptides of the invention may contain one or more modified amino acid residues. The presence of modified amino acids may be advantageous in, for example, increasing polypeptide half-life, reducing polypeptide antigenicity or toxicity, increasing polypeptide storage stability, or the like. Amino acid residue(s) are modified, for example, co-translationally or post-translationally during recombinant production or modified by synthetic or chemical means.

Non-limiting examples of a modified amino acid residue include incorporation or other use of acetylated amino acids, glycosylated amino acids, sulfated amino acids, prenylated (e.g., farnesylated, geranylgeranylated) amino acids, PEG modified (e.g., "PEGylated") amino acids, biotinylated amino acids, carboxylated amino acids, phosphorylated amino acids, etc. References adequate to guide one of skill in the modification of amino acid residues are replete throughout the literature.

The modified amino acid residues may prevent or increase affinity of the polypeptide for another molecule, including, but not limited to, polynucleotide, proteins, carbohydrates, lipids and lipid derivatives, and other organic or synthetic compounds.

XII. Identification of Additional Factors

A transcription factor provided by the present invention can also be used to identify additional endogenous or exogenous molecules that can affect a phenotype or trait of interest. On the one hand, such molecules include organic (small or large molecules) and/or inorganic compounds that affect expression of (i.e., regulate) a particular transcription factor. Alternatively, such molecules include endogenous

molecules that are acted upon either at a transcriptional level by a transcription factor of the invention to modify a phenotype as desired. For example, the transcription factors can be employed to identify one or more downstream gene with which is subject to a regulatory effect of the transcription factor. In one approach, a transcription factor or transcription factor homologue of the invention is expressed in a host cell, e.g., a transgenic plant cell, tissue or explant, and expression products, either RNA or protein, of likely or random targets are monitored, e.g., by hybridization to a microarray of nucleic acid probes corresponding to genes expressed in a tissue or cell type of interest, by two-dimensional gel electrophoresis of protein products, or by any other method known in the art for assessing expression of gene products at the level of RNA or protein. Alternatively, a transcription factor of the invention can be used to identify promoter sequences (i.e., binding sites) involved in the regulation of a downstream target. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence can be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA-binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences can be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the transcription factors with their promoters (Bulyk et al. (1999) Nature Biotechnology 17:573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification can occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or heteropolymer) interactions. Any method suitable for detecting protein-protein interactions can be employed. Among the methods that can be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

The two-hybrid system detects protein interactions in vivo and is described in Chien et al. ((1991), Proc. Natl. Acad. Sci. USA 88:9578-9582) and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain

of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions can be preformed.

XIII. Identification of Modulators

In addition to the intracellular molecules described above, extracellular molecules that alter activity or expression of a transcription factor, either directly or indirectly, can be identified. For example, the methods can entail first placing a candidate molecule in contact with a plant or plant cell. The molecule can be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide can be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence can be detected by use of microarrays, Northern, quantitative PCR, or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1998, and supplements through 2001). Such changes in the expression levels can be correlated with modified plant traits and thus identified molecules can be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

Essentially any available composition can be tested for modulatory activity of expression or activity of any nucleic acid or polypeptide herein. Thus, available libraries of compounds such as chemicals, polypeptides, nucleic acids and the like can be tested for modulatory activity. Often, potential modulator compounds can be dissolved in aqueous or organic (e.g., DMSO-based) solutions for easy delivery to the cell or plant of interest in which the activity of the modulator is to be tested. Optionally, the assays are designed to screen large modulator composition libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays).

In one embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as target compounds.

A combinatorial chemical library can be, e.g., a collection of diverse chemical compounds generated by chemical synthesis or biological synthesis. For example, a combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (e.g., in one example, amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound of a set length). Exemplary libraries include peptide libraries, nucleic acid libraries, antibody libraries (see, e.g., Vaughn et al. (1996) Nature Biotechnology, 14(3):309-314 and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al. Science (1996) 274:1520-1522 and U.S. Patent 5,593,853), peptide nucleic acid libraries (see, e.g., U.S. Patent 5,539,083), and small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337) and the like.

Preparation and screening of combinatorial or other libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Patent 5,010,175; Furka, (1991) Int. J. Pept. Prot. Res. 37:487-493; and Houghton et al. (1991) Nature 354:84-88). Other chemistries for generating chemical diversity libraries can also be used.

In addition, as noted, compound screening equipment for high-throughput screening is generally available, e.g., using any of a number of well known robotic systems that have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations including an automated synthesis apparatus and robotic systems utilizing robotic arms. Any of the above devices are suitable for use with the present invention, e.g., for high-throughput screening of potential modulators. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

Indeed, entire high throughput screening systems are commercially available. These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, microfluidic implementations of screening are also commercially available.

The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. The integrated systems herein, in addition to providing for sequence alignment and, optionally, synthesis of relevant nucleic acids, can include such screening apparatus to identify modulators that have an effect on one or more polynucleotides or polypeptides according to the present invention.

In some assays it is desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. That is, known transcriptional activators or inhibitors can be

incubated with cells/plants/ etc. in one sample of the assay, and the resulting increase/decrease in transcription can be detected by measuring the resulting increase in RNA/ protein expression, etc., according to the methods herein. It will be appreciated that modulators can also be combined with transcriptional activators or inhibitors to find modulators that inhibit transcriptional activation or transcriptional repression. Either expression of the nucleic acids and proteins herein or any additional nucleic acids or proteins activated by the nucleic acids or proteins herein, or both, can be monitored.

In an embodiment, the invention provides a method for identifying compositions that modulate the activity or expression of a polynucleotide or polypeptide of the invention. For example, a test compound, whether a small or large molecule, is placed in contact with a cell, plant (or plant tissue or explant), or composition comprising the polynucleotide or polypeptide of interest and a resulting effect on the cell, plant, (or tissue or explant) or composition is evaluated by monitoring, either directly or indirectly, one or more of: expression level of the polynucleotide or polypeptide, activity (or modulation of the activity) of the polynucleotide or polypeptide. In some cases, an alteration in a plant phenotype can be detected following contact of a plant (or plant cell, or tissue or explant) with the putative modulator, e.g., by modulation of expression or activity of a polynucleotide or polypeptide of the invention. Modulation of expression or activity of a polynucleotide or polypeptide of the invention may also be caused by molecular elements in a signal transduction second messenger pathway and such modulation can affect similar elements in the same or another signal transduction second messenger pathway.

XIV. Subsequences

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least 20, 30, or 50 bases, which hybridize under at least highly stringent (or ultra-high stringent or ultra-ultra-high stringent conditions) conditions to a polynucleotide sequence described above. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted *supra*.

Subsequences of the polynucleotides of the invention, including polynucleotide fragments and oligonucleotides are useful as nucleic acid probes and primers. An oligonucleotide suitable for use as a probe or primer is at least about 15 nucleotides in length, more often at least about 18 nucleotides, often at least about 21 nucleotides, frequently at least about 30 nucleotides, or about 40 nucleotides, or more in length. A nucleic acid probe is useful in hybridization protocols, e.g., to identify additional polypeptide homologues of the invention, including protocols for microarray experiments. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook and Ausubel, *supra*.

In addition, the invention includes an isolated or recombinant polypeptide including a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotides of the invention. For example, such polypeptides, or domains or fragments thereof, can be used as immunogens, e.g., to produce antibodies specific for the polypeptide sequence, or as probes for detecting a sequence of interest. A subsequence can range in size from about 15 amino acids in length up to and including the full length of the polypeptide.

To be encompassed by the present invention, an expressed polypeptide which comprises such a polypeptide subsequence performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA binding domain that binds to a specific DNA promoter region, an activation domain or a domain for protein-protein interactions.

XV. Production of Transgenic Plants

Modification of Traits

The polynucleotides of the invention are favorably employed to produce transgenic plants with various traits, or characteristics, that have been modified in a desirable manner, e.g., to improve the seed characteristics of a plant. For example, alteration of expression levels or patterns (e.g., spatial or temporal expression patterns) of one or more of the transcription factors (or transcription factor homologues) of the invention, as compared with the levels of the same protein found in a wild type plant, can be used to modify a plant's traits. An illustrative example of trait modification, improved characteristics, by altering expression levels of a particular transcription factor is described further in the Examples and the Sequence Listing.

Arabidopsis as a model system

Arabidopsis thaliana is the object of rapidly growing attention as a model for genetics and metabolism in plants. *Arabidopsis* has a small genome, and well documented studies are available. It is easy to grow in large numbers and mutants defining important genetically controlled mechanisms are either available, or can readily be obtained. Various methods to introduce and express isolated homologous genes are available (see Koncz, et al., eds. *Methods in Arabidopsis Research*. et al. (1992), World Scientific, New Jersey, New Jersey, in "Preface"). Because of its small size, short life cycle, obligate autogamy and high fertility, *Arabidopsis* is also a choice organism for the isolation of mutants and studies in morphogenetic and development pathways, and control of these pathways by transcription factors (Koncz, *supra*, p. 72). A number of studies introducing transcription factors into *A. thaliana* have demonstrated the utility of this plant for understanding the mechanisms of gene regulation and trait alteration in plants. See, for example, Koncz, *supra*, and U.S. Patent Number 6,417,428).

Arabidopsis genes in transgenic plants.

Expression of genes which encode transcription factors modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription factors may also modify expression of endogenous genes, polynucleotides, and

proteins. Examples include Peng et al. (1997, *Genes and Development* 11:3194-3205) and Peng et al. (1999, *Nature*, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, *Plant Cell* 13:1791-1802); Nandi et al. (2000, *Curr. Biol.* 10:215-218); Coupland (1995, *Nature* 377:482-483); and Weigel and Nilsson (1995, *Nature* 377:482-500).

Homologous genes introduced into transgenic plants.

Homologous genes that may be derived from any plant, or from any source whether natural, synthetic, semi-synthetic or recombinant, and that share significant sequence identity or similarity to those provided by the present invention, may be introduced into plants, for example, crop plants, to confer desirable or improved traits. Consequently, transgenic plants may be produced that comprise a recombinant expression vector or cassette with a promoter operably linked to one or more sequences homologous to presently disclosed sequences. The promoter may be, for example, a plant or viral promoter.

The invention thus provides for methods for preparing transgenic plants, and for modifying plant traits. These methods include introducing into a plant a recombinant expression vector or cassette comprising a functional promoter operably linked to one or more sequences homologous to presently disclosed sequences. Plants and kits for producing these plants that result from the application of these methods are also encompassed by the present invention.

The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4, Table 5, and Table 6.

Table 4

Polynucleotide SEQ ID NO:	GID No.	Trait	Category	Family	Comment	Polypeptide SEQ ID NO:	Conserved domains
1	G1275	Architecture; size	Dev and morph	WRKY	Reduced apical dominance; small plant	2	(113-169)
3	G1411	Architecture	Dev and morph	AP2	Loss of apical dominance	4	(87-154)
5	G1488	Architecture; light response; size; seed protein content	Dev and morph; seed biochemistry	GATA/Zn	Reduced apical dominance, shorter stems; constitutive photomorphogenesis; reduced size; altered seed protein content	6	(221-246)
7	G1499	Architecture; flower; morphology; other	Dev and morph	HLH/MYC	Altered plant architecture; altered floral organ identity and development; dark green color	8	(118-181)
9	G1543	Architecture; flower; morphology; other; seed oil	Dev and morph; seed biochemistry	HB	Altered plant architecture; altered carpel shape; dark green color; decreased seed oil	10	(135-195)
11	G1635	Architecture; morphology; other; fertility	Dev and morph	MYB-related	Reduced apical dominance; pale green, smaller plants; reduced fertility	12	(44-104)
13	G1794	Architecture; light response; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Altered plant architecture; constitutive photomorphogenesis; altered seed oil and protein content	14	(182-248)
15	G1839	Architecture; size	Dev and morph	AP2	Altered plant architecture; reduced size	16	(118-184)
17	G2108	Architecture	Dev and morph	AP2	Altered inflorescence structure	18	(18-85)
19	G2291	Architecture; flowering time	Dev and morph; flowering time	AP2	Altered plant architecture; late flowering	20	(TBD)
21	G2452	Architecture; leaf	Dev and morph	MYB-related	Reduced apical dominance; pale green color	22	(27-213)
23	G2509	Architecture; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced apical dominance; altered seed oil and protein content	24	(89-156)
25	G390	Architecture	Dev and morph	HB	Altered shoot development	26	(18-81)
27	G391	Architecture	Dev and morph	HB	Altered shoot development	28	(25-85)
29	G438	Architecture; stem	Dev and morph	HB	Reduced branching; reduced lignin	30	(22-85)

Table 4

31	G47	Architecture; stem; flowering time; altered seed oil content	Dev and morph; flowering time; seed biochemistry	AP2	Altered architecture and inflorescence development, structure of vascular tissues; late flowering; altered seed oil content	32	(11-80)
33	G559	Architecture; fertility	Dev and morph	bZIP	Loss of apical dominance; reduced fertility	34	(203-264)
35	G568	Architecture; flowering time	Dev and morph	bZIP	Altered branching; late flowering	36	(215-265)
37	G580	Architecture; flower	Dev and morph	bZIP	Altered inflorescences; altered flower development	38	(162-218)
39	G615	Architecture; fertility	Dev and morph	TEO	Altered plant architecture; little or no pollen production, poor filament elongation	40	(88-147)
41	G732	Architecture; flower; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Reduced apical dominance; abnormal flowers; altered seed oil and protein content	42	(31-91)
43	G988	Architecture; fertility; flower; stem; seed oil and protein content	Dev and morph; seed biochemistry	SCR	Reduced lateral branching; reduced fertility; enlarged floral organs, short pedicels; thicker stem, altered distribution of vascular bundles; altered seed oil and protein content	44	(178-195)
45	G1519	Embryo lethal	Dev and morph	RING/C3HC4	Embryo lethal	46	(327-364)
47	G374	Embryo lethal	Dev and morph	Z-ZPF	Embryo lethal	48	(35-67, 245-277)
49	G877	Embryo lethal	Dev and morph	WRKY	Embryo lethal	50	(272-328, 487-603)
51	G1000	Fertility; size; flower; stem	Dev and morph	MYB-(R1)R2R3	Reduced fertility; small plant; reduced or absent petals and sepals; reduced inflorescence, stem elongation	52	(14-117)
53	G1067	Fertility; leaf; size	Dev and morph	AT-hook	Reduced fertility; altered leaf shape; small plant	54	(86-93)
55	G1075	Fertility; flower; leaf; size	Dev and morph	AT-hook	Reduced fertility; reduced or absent petals, sepals and stamens; altered leaf shape; small plant	56	(78-85)
57	G1266	Fertility; size	Dev and morph	AP2	Reduced fertility; small plant	58	(79-147)
59	G1311	Fertility; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; small plant	60	(11-112)
61	G1321	Fertility; flower	Dev and morph	MYB-(R1)R2R3	Poor fertility; altered flower morphology	62	(4-106)
63	G1326	Fertility; flower; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; petals and sepals are smaller; small plant	64	(18-121)
65	G1367	Fertility; size	Dev and morph	AT-hook	Reduced fertility; reduced size	66	(179-201, 262-285, 298-319, 335-357)
67	G1386	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; reduced size; altered seed oil and protein content	68	(TBD)

Table 4

69	G1421	Fertility; size; seed oil content	Dev and morph; seed biochemistry	AP2	Reduced fertility; small plant; altered seed oil content	70	(74-151)
71	G1453	Fertility; morphology: other			Reduced fertility; altered inflorescence development	72	(13-160)
73	G1560	Fertility; flower; size	Dev and morph	NAC	Reduced fertility; altered flower development; reduced size	74	(62-151)
75	G1594	Fertility; leaf; seed	Dev and morph	HS	Reduced fertility; altered leaf shape and development; large pale seed	76	(343-308)
77	G1750	Fertility; size; seed oil content	Dev and morph; seed biochemistry	HB	Reduced fertility; reduced size; increased seed oil content	78	(107-173)
79	G1947	Fertility; flower; seed protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; extended period of flowering; altered seed protein content	80	(37-120)
81	G2011	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	HS	Reduced fertility; reduced size; altered seed oil and protein content	82	(56-147)
83	G2094	Fertility; leaf; size	Dev and morph	GATA/Zn	Reduced fertility; altered leaf development; reduced size	84	(43-68)
85	G2113	Fertility; leaf; seed protein content	Dev and morph; seed biochemistry		Reduced fertility; long petioles, altered orientation; altered seed protein content	86	(TBD)
87	G2115	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	88	(46-115)
89	G2130	Fertility; size; senescence	Dev and morph	AP2	Reduced fertility; reduced size; early senescence	90	(93-160)
91	G2147	Fertility; size	Dev and morph	HLH/MYC	Reduced fertility; reduced size	92	(160-234)
93	G2156	Fertility; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced fertility; reduced size; altered seed protein content	94	(66-86)
95	G2294	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	96	(32-102)
97	G2510	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	98	(41-108)
99	G2893	Fertility; flower; size	Dev and morph	MYB-(R1)R2R3	Reduced fertility; altered flower development; reduced size	100	(19-120)
101	G340	Fertility; size	Dev and morph	Z-C3H	Reduced fertility; size	102	(37-154)
103	G39	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	104	(24-90)
105	G439	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	106	(110-177)
107	G470	Fertility	Dev and morph	ARF	Short stamen filaments	108	(61-393)

Table 4

109	G652	Fertility; seed; flower; size; seed oil content	Dev and morph; seed biochemistry	Z-CLDSH	Reduced fertility; irregular shaped seed; altered flower development; reduced size, slow growth; altered seed oil content	110	(28-49, 137-151, 182-196)
111	G671	Fertility; flower; leaf; size; stem	Dev and morph; biochemistry	MYB-(R1)R2R3	Reduced fertility; reduced petal abscission; altered leaf shape; small plant; altered inflorescence stem structure	112	(15-115)
113	G779	Fertility; flower	Dev and morph; biochemistry	HLH/MYC	Reduced fertility, homeotic transformations	114	(126-182)
115	G962	Fertility; size	Dev and morph; biochemistry	NAC	Reduced fertility; small plant	116	(53-175)
117	G977	Fertility; leaf; morphology; other; size	Dev and morph; biochemistry	AP2	Reduced fertility; altered leaf shape; dark green; small plant	118	(5-72)
119	G1063	Flower; leaf; inflorescence; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development; altered seed oil and protein content	120	(131-182)
121	G1140	Flower	Dev and morph; biochemistry	MADS	Altered flower development	122	(2-57)
123	G1425	Flower	Dev and morph; biochemistry	NAC	Altered flower and inflorescence development	124	(20-173)
125	G1449	Flower	Dev and morph; biochemistry	IAA	Altered flower structure	126	(48-53, 74-107, 122-152)
127	G1897	Flower; leaf; seed protein content	Dev and morph; seed biochemistry	Z-Dof	Altered flower development; altered leaf development; altered seed protein content	128	(34-62)
129	G2143	Flower; leaf; inflorescence	Dev and morph; biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development	130	(128-179)
131	G2535	Flower; seed protein content	Dev and morph; seed biochemistry	NAC	Altered flower development; altered seed protein content	132	(11-114)
133	G2557	Flower; leaf	Dev and morph; biochemistry	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color	134	(278-328)
135	G259	Flower; leaf	Dev and morph; biochemistry	HS	Altered flower development; altered leaf development	136	(27-131)
137	G353	Flower; leaf; size; seed protein content	Dev and morph; seed biochemistry	Z-C2H2	Short pedicels, downward pointing siliques; altered leaf development; reduced size; altered seed protein content	138	(41-61, 84-104)
139	G354	Flower; light response; size	Dev and morph; biochemistry	Z-C2H2	Short pedicels, downward pointing siliques; constitutive morphogenesis; reduced size	140	(42-62, 88-109)
141	G638	Flower; morphology; other	Dev and morph; biochemistry	TH	Altered flower development; multiple developmental defects	142	(119-206)

Table 4

143	G869	Flower; morphology; other; seed oil	Dev and morph; seed biochemistry	AP2	Abnormal anther development; altered seed fatty acids	144	(109-177)
145	G1645	Inflorescence; leaf	Dev and morph	MYB-(R1)R2R3	Altered inflorescence structure; altered leaf development	146	(90-210)
147	G1038	Leaf	Dev and morph	GARP	Altered leaf shape	148	(198-247)
149	G1073	Leaf; size; flowering time	Dev and morph; flowering time	AT-hook	Serrated leaves; increased plant size; flowering appears to be slightly delayed	150	(33-42, 78-175)
151	G1146	Leaf	Dev and morph	PAZ	Altered leaf development	152	(886-896)
153	G1267	Leaf; size	Dev and morph	WRKY	Dark green shiny leaves; small plant	154	(70-127)
155	G1269	Leaf	Dev and morph	MYB-related	Long petioles, upturned leaves	156	(27-83)
157	G1452	Leaf; trichome; flowering time	Dev and morph; flowering time	NAC	Altered leaf shape, dark green color; reduced trichome density; late flowering	158	(30-177)
159	G1494	Leaf; size; light response; seed	Dev and morph	HLH/MYC	Pale green leaves, altered leaf shape; reduced size; long hypocotyls; large, pale seeds	160	(261-311)
161	G1548	Leaf	Dev and morph	HB	Altered leaf development	162	(17-77)
163	G1574	Leaf	Dev and morph	SWI/SNF	Altered leaf development	164	(28-350)
165	G1586	Leaf; size	Dev and morph	HB	Narrow leaves; small plants	166	(21-81)
167	G1786	Leaf; light response; size	Dev and morph	MYB-(R1)R2R3	Dark green, small leaves with short petioles; photomorphogenesis in the dark; small plant	168	(TBD)
169	G1792	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Dark green, shiny leaves; altered seed oil and protein content	170	(17-85)
171	G1865	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	GRF-like	Altered leaf development; altered seed oil and protein content	172	(124-149)
173	G1886	Leaf; size	Dev and morph	Z-Dof	Chlorotic patches in leaves; reduced size	174	(17-59)
175	G1933	Leaf; size; seed protein content	Dev and morph; seed biochemistry	WRKY	Altered leaf development; reduced size; altered seed protein content	176	(205-263, 344-404)
177	G2059	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Smaller, curled leaves; altered seed oil, protein content	178	(184-254)
179	G2105	Leaf; seed	Dev and morph	TH	Alterations in leaf surface; large, pale seeds	180	(100-153)
181	G2117	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	bZIP	Small, dark green leaves; altered seed oil and protein content	182	(46-106)

Table 4

183	G2124	Leaf; seed protein content	Dev and morph; seed biochemistry	TEO	Altered leaf development; altered seed protein content	184	(75-132)
185	G2140	Leaf; root	Dev and morph	HLH/MYC	Altered leaf development; short roots	186	(167-242)
187	G2144	Leaf; light response; size; seed oil content	Dev and morph; seed biochemistry	HLH/MYC	Pale green leaves, altered leaf shape; long hypocotyls; reduced size; altered seed oil content	188	(203-283)
189	G2431	Leaf	Dev and morph	GARP	Dark green leaves; reduced size	190	(38-88)
191	G2465	Morphology: other; leaf	Dev and morph	GARP	Slowed development; altered leaf color and shape	192	(219-269)
193	G2583	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Glossy, shiny leaves; altered seed oil and protein content	194	(4-71)
195	G2724	Leaf	Dev and morph	MYB-(R1)R2R3	Dark green leaves	196	(7-113)
197	G377	Leaf; morphology: other	Dev and morph	RING/C3H2C3	Altered leaf development; slow growth	198	(85-128)
199	G428	Leaf	Dev and morph	HB	Altered leaf shape	200	(229-292)
201	G447	Leaf; morphology: other; size	Dev and morph	ARF	Dark green leaves; altered cotyledon shape; reduced size	202	(22-356)
203	G464	Leaf	Dev and morph	IAA	Altered leaf shape	204	(20-28, 71-82, 126-142, 187-224)
205	G557	Leaf; size	Dev and morph	bZIP	Dark green color; small plant	206	(90-150)
207	G577	Leaf	Dev and morph	BZIP2	Reduced size, increased anthocyanins	208	(TBD)
209	G674	Leaf; size	Dev and morph	MYB-(R1)R2R3	Dark green leaves, upwardly oriented; reduced size	210	(20-120)
211	G736	Leaf; flowering time	Dev and morph; flowering time	Z-Dof	Altered leaf shape; later flowering	212	(54-111)
213	G903	Leaf	Dev and morph	Z-C2H2	Altered leaf morphology	214	(68-92)
215	G917	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	MADS	Altered leaf development; altered seed oil and protein content	216	(2-57)
217	G921	Leaf	Dev and morph	WRKY	Serrated leaves	218	(146-203)
219	G922	Leaf; size	Dev and morph	SCR	Altered development, dark green color; reduced size	220	(225-242)
221	G932	Leaf; size	Dev and morph	MYB-(R1)R2R3	Altered development, dark green color; reduced size	222	(12-118)
223	G599	Leaf; size	Dev and morph	DBP	Altered leaf shape; small plant	224	(187-219, 264-300)
225	G804	Leaf; size	Dev and morph	PCF	Altered leaf shape, small plant	226	(54-117)

Table 4

227	G1062	Light response; morphology; other; seed	Dev and morph	HLH/MYC	Constitutive photomorphogenesis; slow growth; altered seed shape	228	(308-359)
229	G1322	Light response; size	Dev and morph	MYB-(R1)R2R3	Photomorphogenesis in the dark; reduced size	230	(26-130)
231	G1331	Light response; morphology; other; seed oil and protein content	Dev and morph; seed biochemistry		Constitutive photomorphogenesis; multiple developmental alterations; altered seed oil and protein content	232	(8-109)
233	G1521	Light response	Dev and morph	MYB-(R1)R2R3 RING/C3HC4	Constitutive photomorphogenesis	234	(39-80)
235	G183	Light response; seed protein content	Dev and morph; seed biochemistry	WRKY	Constitutive photomorphogenesis; altered seed protein content	236	(307-363)
237	G2555	Light response	Dev and morph	HLH/MYC	Constitutive photomorphogenesis	238	(175-245)
239	G375	Light response	Dev and morph	Z-Dof	Upward pointing leaves	240	(75-103)
241	G1007	Morphology; other	Dev and morph	AP2	Multiple developmental alterations	242	(TBD)
243	G1010	Morphology; other	Dev and morph	ABI3/VP-1	Multiple developmental alterations	244	(33-122)
245	G1014	Morphology; other; trichome	Dev and morph	ABI3/VP-1	Multiple developmental defects; reduced trichomes	246	(90-172)
247	G1035	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	248	(39-91)
249	G1046	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	250	(79-138)
251	G1049	Morphology; other; seed protein content	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; altered seed protein content	252	(77-132)
253	G1069	Morphology; other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	254	(67-74)
255	G1070	Morphology; other	Dev and morph	AT-hook	Several developmental defects	256	(98-120)
257	G1076	Morphology; other	Dev and morph	AT-hook	Lethal when overexpressed	258	(82-89)
259	G1089	Morphology; other	Dev and morph	BZIP2	Developmental defects at seedling stage	260	(425-500)
261	G1093	Morphology; other	Dev and morph	RING/C3H2C3	Multiple morphological alterations	262	(105-148)

Table 4

263	G1127	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	264	(103-110, 155-162)
265	G1131	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed protein content	266	(173-220)
267	G1145	Morphology: other; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; reduced seed size, altered seed shape; altered seed oil and protein content	268	(227-270)
269	G1229	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Several developmental defects; altered seed oil and protein content	270	(102-160)
271	G1246	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	272	(27-139)
273	G1255	Morphology: other; seed	Dev and morph	Z-CO-like	Reduced apical dominance; increased seed size	274	(18-56)
275	G1304	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	276	(13-118)
277	G1318	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	278	(20-123)
279	G1320	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	280	(5-108)
281	G1330	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	282	(28-134)
283	G1352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	284	(108-129, 167-188)
285	G1354	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	286	(TBD)
287	G1360	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	288	(18-174)
289	G1364	Morphology: other	Dev and morph	CAAT	Lethal when overexpressed	290	(29-120)
291	G1379	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	292	(18-85)
293	G1384	Morphology: other	Dev and morph	AP2	Abnormal inflorescence and flower development	294	(TBD)
295	G1399	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	296	(86-93)

Table 4

297	G1415	Morphology: other	Dev and morph morph; seed biochemistry	AP2	Multiple developmental alterations	298	(TBD)
299	G1417	Morphology: other; seed oil		WRKY	Reduced seedling germination and vigor; increase in 18:2, decrease in 18:3	300	(239-296)
301	G1442	Morphology: other	Dev and morph	GRF-like	Multiple developmental alterations	302	(172-223)
303	G1454	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed oil and protein content	304	(9-178)
305	G1459	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	306	(10-152)
307	G1460	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	308	(TBD)
309	G147	Morphology: other	Dev and morph	MADS	Multiple developmental defects	310	(2-57)
311	G1471	Morphology: other; seed oil	Dev and morph; seed biochemistry	Z-C2H2	Multiple developmental alterations; increased seed oil content	312	(49-70)
313	G1475	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	314	(51-73)
315	G1477	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	316	(29-48)
317	G1487	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	318	(251-276)
319	G1492	Morphology: other	Dev and morph	GARP	Multiple developmental alterations	320	(34-83)
321	G1531	Morphology: other; seed; seed protein content	Dev and morph; seed biochemistry	RING/C3HC4	Multiple developmental alterations; pale seed; altered seed protein content	322	(41-77)
323	G1540	Morphology: other	Dev and morph	HB	Reduced cell differentiation in meristem	324	(35-98)
325	G1544	Morphology: other	Dev and morph	HB	Multiple developmental alterations	326	(64-124)

Table 4

327	G156	Morphology: other; seed	Dev and morph	MADS	Multiple developmental defects; seed color alteration	328	(2-57)
329	G1584	Morphology: other	Dev and morph	HB	Multiple developmental alterations	330	(TBD)
331	G1587	Morphology: other	Dev and morph	HB	Multiple developmental alterations	332	(61-121)
333	G1588	Morphology: other	Dev and morph	HB	Multiple developmental alterations	334	(66-124)
335	G1589	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HB	Multiple developmental alterations; altered seed protein content	336	(384-448)
337	G160	Morphology: other	Dev and morph	MADS	Multiple developmental defects	338	(7-62)
339	G1636	Morphology: other	Dev and morph	MYB-related	Pale green, smaller plants	340	(100-165)
341	G1642	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	342	(TBD)
343	G1747	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	344	(11-114)
345	G1749	Morphology: other	Dev and morph	AP2	Multiple developmental alterations; formation of necrotic lesions	346	(84-155)
347	G1751	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	348	(TBD)
349	G1752	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	350	(83-151)
351	G1763	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	352	(140-209)
353	G1766	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	354	(10-153)
355	G1767	Morphology: other; seed oil content	Dev and morph; seed biochemistry	SCR	Multiple developmental alterations; altered seed oil content	356	(255-272)
357	G1778	Morphology: other	Dev and morph	GATA/Zn	Lethal when overexpressed	358	(94-119)
359	G1789	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-related	Delayed development; altered seed protein content	360	(1-50)
361	G1790	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	362	(217-316)

Table 4

363	G1791	Morphology: other	Dev and morph Dev and	AP2	Multiple developmental alterations	364	(TBD)
365	G1793	Morphology: other; seed oil	morph; seed biochemistry	AP2	Multiple developmental alterations; increased seed oil content	366	(179-255, 281-349)
367	G1795	Morphology: other; trichome	Dev and morph	AP2	Multiple developmental alterations; reduced trichomes	368	(12-80)
369	G1800	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	370	(TBD)
371	G1806	Morphology: other	Dev and morph	bZIP	Multiple developmental alterations	372	(165-225)
373	G1811	Morphology: other	Dev and morph	ABI3/VP-1	Multiple developmental alterations	374	(TBD)
375	G182	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	376	(217-276)
377	G1835	Morphology: other	Dev and morph	GATA/Zn	Small, spindly plant	378	(224-296)
379	G1836	Morphology: other	Dev and morph	CAAT	Pale green	380	(30-164)
381	G1838	Morphology: other; seed oil content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; increased seed oil content	382	(229-305, 330-400)
383	G1843	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	384	(2-57)
385	G1853	Morphology: other	Dev and morph	AKR	Lethal when overexpressed	386	(entire protein)
387	G1855	Morphology: other	Dev and morph	AKR	Slow growth	388	(entire protein)
389	G187	Morphology: other	Dev and morph	WRKY	Variety of morphological alterations	390	(172-228)
391	G1881	Morphology: other	Dev and morph	Z-CO-like	Multiple developmental alterations	392	(5-28, 56-79)
393	G1882	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	394	(97-125)
395	G1883	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	396	(82-124)
397	G1884	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	398	(43-71)
399	G1891	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	400	(27-69)

Table 4

401	G1896	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	402	(43-85)
403	G1898	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	404	(31-59)
405	G1902	Morphology: other; seed oil content	Dev and morph; seed biochemistry	Z-Dof	Multiple developmental alterations; increased seed oil content	406	(31-59)
407	G1904	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	408	(53-95)
409	G1906	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	410	(19-47)
411	G1913	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	412	(27-55)
413	G1914	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	414	(195-216, 245-266)
415	G1925	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	416	(6-150)
417	G1929	Morphology: other	Dev and morph	Z-CO-like	Slow growth, delayed development	418	(31-53)
419	G1930	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	420	(59-124)
421	G195	Morphology: other	Dev and morph	WRKY	Multiple developmental defects	422	(183-239)
423	G1954	Morphology: other	Dev and morph	HLH/MYC	Lethal when overexpressed	424	(187-259)
425	G1958	Morphology: other; seed protein content	Dev and morph; seed biochemistry	GARP	Reduced size and root mass in plates; altered seed protein content	426	(230-278)
427	G196	Morphology: other; seed protein content	Dev and morph; seed biochemistry	WRKY	Multiple developmental alterations; altered seed protein content	428	(223-283)
429	G1965	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	430	(27-55)
431	G1976	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	432	(219-323)
433	G2057	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	434	(TBD)
435	G2107	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	436	(TBD)

Table 4

437	G211	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	438	(24-137)
439	G2133	Morphology: other; flowering time; seed protein content	Dev and morph; flowering time	AP2	Multiple developmental alterations; late flowering; altered seed protein content	440	(11-83)
441	G2134	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	442	(TBD)
443	G2151	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil and protein content	444	(93-113, 124-144)
445	G2154	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	446	(97-119)
447	G2157	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	448	(82-102, 164-107)
449	G2181	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	450	(22-169)
451	G221	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	452	(21-125)
453	G2290	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	454	(147-205)
455	G2299	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	456	(48-115)
457	G2340	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Tissue necrosis; multiple developmental alterations; altered seed oil and protein content	458	(14-120)
459	G2346	Morphology: other	Dev and morph	SBP	Enlarged seedlings	460	(59-135)
461	G237	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	462	(11-113)
463	G2373	Morphology: other; seed protein content	Dev and morph; seed biochemistry	TH	Multiple developmental alterations; altered seed protein content	464	(290-350)
465	G2376	Morphology: other; seed oil protein	Dev and morph; seed biochemistry	TH	Seedling lethality; altered seed protein content	466	(79-178, 336-408)
467	G24	Morphology: other	Dev and morph	AP2	Reduced size and necrotic patches	468	(25-93)

Table 4

469	G2424	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	470	(107-219)
471	G2505	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	472	(10-159)
473	G2512	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	474	(79-139)
475	G2513	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	476	(TBD)
477	G2519	Morphology: other	Dev and morph	HLH/MYC	Multiple developmental alterations	478	(1-65)
479	G2520	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed oil and protein content	480	(135-206)
481	G2533	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	482	(11-186)
483	G2534	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	484	(10-157)
485	G2573	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed oil and protein content	486	(31-98)
487	G2589	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	488	(2-57)
489	G2687	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	490	(51-120)
491	G27	Morphology: other	Dev and morph	AP2	Abnormal development, small	492	(37-104)
493	G2720	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed oil and protein content	494	(10-114)
495	G2787	Morphology: other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	496	(172-192, 226-247, 256-276, 290-311, 245 366)
497	G2789	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	498	(53-73, 121-165)
499	G31	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	500	(TBD)

Table 4

501	G33	Morphology: other	Dev and morph	AP2	Multiple developmental defects	502	(50-117)
503	G342	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	504	(155-190)
505	G352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	506	(99-119, 166-186)
507	G357	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	508	(7-29)
509	G358	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	510	(124-135, 188-210)
511	G360	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	512	(42-62)
513	G362	Size; Morphology: other; trichome; flowering time; seed protein content	Dev and morph; flowering time; seed biochemistry	Z-C2H2	Reduced size; increased pigmentation in seed, embryos and other organs; ectopic trichome formation; increased trichome number; late flowering; altered protein content	514	(62-82)
515	G364	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	516	(54-76)
517	G365	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	518	(70-90)
519	G367	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	520	(63-84)
521	G373	Morphology: other	Dev and morph	RING/C3HC4	Multiple developmental alterations	522	(129-168)
523	G396	Morphology: other; size	Dev and morph	HB	Altered leaf coloration and shape, reduced fertility; small plant	524	(159-220)
525	G431	Morphology: other	Dev and morph	HB	Developmental defect, sterile	526	(286-335)
527	G479	Morphology: other	Dev and morph	SBP	Multiple developmental alterations	528	(70-149)
529	G546	Morphology: other	Dev and morph	RING/C3H2C3	Slow growth and development; increased anthocyanin pigmentation	530	(114-155)
531	G551	Morphology: other	Dev and morph	HB	Multiple developmental alterations	532	(73-133)
533	G578	Morphology: other	Dev and morph	bZIP	Lethal when overexpressed	534	(36-96)

Table 4

535	G596	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	536	(89-96)
537	G617	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	538	(64-118)
539	G620	Morphology: other; seed protein content	Dev and morph; seed biochemistry	CAAT	Multiple developmental alterations; altered seed protein content	540	(20-118)
541	G625	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	542	(52-119)
543	G658	Morphology: other	Dev and morph	MYB-(R1)R2R3	Developmental defect	544	(2-105)
545	G716	Morphology: other	Dev and morph	ARF	Multiple developmental defects	546	(24-355)
547	G725	Morphology: other	Dev and morph	GARP	Developmental defect	548	(39-87)
549	G727	Morphology: other	Dev and morph	GARP	Multiple morphological alterations	550	(226-269)
551	G740	Morphology: other	Dev and morph	Z-CLDSH	Slow growth	552	(24-42, 232-268)
553	G770	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	554	(19-162)
555	G858	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	556	(2-57)
557	G865	Morphology: other; seed protein content	Dev and morph; seed biochemistry	AP2	Altered morphology; increased seed protein	558	(36-103)
559	G872	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	560	(18-85)
561	G904	Morphology: other	Dev and morph	RING/C3H2C3	Multiple developmental alterations	562	(117-158)
563	G910	Morphology: other; flowering time	Dev and morph; flowering time	Z-CO-like	Multiple developmental alterations; late flowering	564	(14-37, 77-103)
565	G912	Morphology: other; size; sugar sensing; flowering time	Dev and morph; sugar sensing; flowering time	AP2	Dark green color; small plant; reduced cotyledon expansion in glucose; late flowering	566	(51-118)
567	G920	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	568	(152-211)

Table 4

569	G939	Morphology: other; size	Dev and morph Dev and morph; seed biochemistry	EIL	Pale seedlings on agar; reduced size	570	(97-106)
571	G963	Morphology: other; seed protein content		NAC	Slowed growth rate; altered seed protein content	572	(TBD)
573	G979	Morphology: other; seed	Dev and morph	AP2	Several developmental defects; altered seed development, ripening and germination	574	(63-139, 165-233)
575	G987	Morphology: other	Dev and morph	SCR	Developmental defects	576	(428-432, 704-708)
577	G993	Morphology: other; seed protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed protein content	578	(69-134)
579	G681	Morphology: other; leaf glucosinolates	Dev and morph; leaf biochemistry		Multiple developmental alterations; overexpression results in an increase in M39480	580	(14-120)
581	G1482	Root	Dev and morph	MYB-(R1)R2R3 Z-CO-like	Increased root growth	582	(5-63)
583	G225	Root; trichome	Dev and morph	MYB-related	Increased root hairs; glabrous, lack of trichomes	584	(39-76)
585	G226	Root; trichome; seed protein content	Dev and morph; seed biochemistry	MYB-related	Increased root hairs; glabrous, lack of trichomes; increased seed protein	586	(28-78)
587	G9	Root	Dev and morph	AP2	Increased root mass	588	(62-127)
589	G1040	Seed	Dev and morph	GARP	Smaller and more rounded seeds	590	(109-158)
591	G2114	Seed	Dev and morph	AP2	Increased seed size	592	(221-297, 323-393)
593	G450	Seed; size; seed protein content	Dev and morph; seed biochemistry	IAA	Increased seed size; reduced plant size; altered seed protein content	594	(TBD)
595	G584	Seed	Dev and morph	HLH/MYC	Large seeds	596	(401-494)
597	G668	Seed	Dev and morph	MYB-(R1)R2R3	Reduced seed color	598	(13-113)
599	G1050	Senescence	Dev and morph	bZIP	Delayed senescence	600	(372-425)
601	G1463	Senescence	Dev and morph	NAC	Premature senescence	602	(9-156)
603	G1944	Senescence; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Early senescence; reduced size; altered seed protein content	604	(87-100)
605	G2383	Senescence; seed protein content	Dev and morph; seed biochemistry	TEO	Early senescence; altered seed protein content	606	(89-149)
607	G571	Senescence; flowering time	Dev and morph; flowering time	bZIP	Delayed senescence; late flowering	608	(160-220)

Table 4

609	G636	Senescence; size	Dev and morph Dev and morph; flowering time	TH	Premature senescence; reduced size	610	(55-145, 405-498)
611	G878	Senescence; flowering time	flowering time	WRKY	Delayed senescence; late flowering	612	(250-305, 415-475)
613	G1134	Silique	Dev and morph	HLH/MYC	Siliques with altered shape	614	(198-247)
615	G1008	Size	Dev and morph	AP2	Small plant	616	(96-163)
617	G1020	Size	Dev and morph	AP2	Very small T1 plants	618	(28-95)
619	G1023	Size	Dev and morph	AP2	Reduced size	620	(128-195)
621	G1053	Size	Dev and morph	bZIP	Small plant	622	(74-120)
623	G1137	Size	Dev and morph	HLH/MYC	Small T1 plants	624	(264-314)
625	G1181	Size	Dev and morph	HS	Small T1 plants	626	(24-114)
627	G1228	Size	Dev and morph	HLH/MYC	Reduced size	628	(179-233)
629	G1277	Size	Dev and morph	AP2	Small plant	630	(18-85)
631	G1309	Size	Dev and morph	MYB-(R1)R2R3	Small plant	632	(9-114)
633	G1314	Size; sugar sensing; seed protein content	Dev and morph; sugar sensing; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced seedling vigor on high glucose; altered seed protein content	634	(14-116)
635	G1317	Size	Dev and morph	MYB-(R1)R2R3	Reduced size	636	(13-118)
637	G1323	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Small T1 plants, dark green; decreased seed oil, increased seed protein	638	(15-116)
639	G1332	Size; trichome; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced trichome density; altered seed oil and protein content	640	(13-116)
641	G1334	Size	Dev and morph	CAAT	Small, dark green	642	(18-190)
643	G1381	Size	Dev and morph	AP2	Reduced size	644	(68-135)
645	G1382	Size	Dev and morph	WRKY	Small plant	646	(210-266, 385-437)
647	G1435	Size; flowering time	Dev and morph; flowering time	GARP	Increased plant size; late flowering	648	(146-194)
649	G1537	Size	Dev and morph	HB	Small T1 plants with altered development	650	(14-74)
651	G1545	Size	Dev and morph	HB	Reduced size	652	(54-117)
653	G1641	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-related	Small plant; altered seed oil and protein content	654	(139-200)
655	G165	Size; seed protein content	Dev and morph; seed biochemistry	MADS	Reduced size; altered seed protein content	656	(7-62)

Table 4

657	G1652	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	658	(143-215)
659	G1655	Size	Dev and morph	HLH/MYC	Small plant	660	(134-192)
661	G1671	Size	Dev and morph	NAC	Reduced size	662	(TBD)
663	G1756	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Reduced size; altered seed protein content	664	(TBD)
665	G1757	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Small plant; altered seed protein content	666	(158-218)
667	G1782	Size	Dev and morph	CAAT	Small, spindly plant	668	(166-238)
669	G184	Size	Dev and morph	WRKY	Small plant	670	(295-352)
671	G1845	Size	Dev and morph	AP2	Small plant	672	(140-207)
673	G1879	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	674	(107-176)
675	G1888	Size	Dev and morph	Z-CO-like	Reduced size, dark green leaves	676	(5-50)
677	G189	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Increased leaf size; altered seed protein content	678	(240-297)
679	G1939	Size	Dev and morph	PCF	Reduced size	680	(40-102)
681	G194	Size	Dev and morph	WRKY	Small plant	682	(174-230)
683	G1943	Size	Dev and morph	HLH/MYC	Reduced size	684	(335-406)
685	G21	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	686	(97-164)
687	G2132	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	688	(TBD)
689	G2145	Size	Dev and morph	HLH/MYC	Reduced size	690	(166-243)
691	G23	Size	Dev and morph	AP2	Small T1 plants	692	(61-117)
693	G2313	Size	Dev and morph	MYB-related	Reduced size	694	(TBD)
695	G2344	Size	Dev and morph	CAAT	Reduced size, slow growth	696	(TBD)
697	G2430	Size	Dev and morph	GARP	Increased leaf size, faster development	698	(425-478)
699	G2517	Size	Dev and morph	WRKY	Reduced size	700	(118-234)
701	G2521	Size	Dev and morph	HLH/MYC	Reduced size	702	(145-213)
703	G258	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; altered seed oil and protein content	704	(24-124)

Table 4

705	G280	Size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced size; altered seed protein content	706	(97-104, 130-137-155-162, 185-192)
707	G3	Size	Dev and morph	AP2	Small plant	708	(28-95)
709	G343	Size	Dev and morph	GATA/Zn	Small plant	710	(178-214)
711	G363	Size	Dev and morph	Z-C2H2	Small plant	712	(87-108)
713	G370	Size	Dev and morph	Z-C2H2	Reduced size, shiny leaves	714	(97-117)
715	G385	Size	Dev and morph	HB	Small plant, short inflorescence stems, dark green	716	(60-123)
717	G439	Size	Dev and morph	AP2	Small plant	718	(110-177)
719	G440	Size	Dev and morph	AP2	Small plant	720	(122-189)
721	G5	Size	Dev and morph	AP2	Small plant	722	(149-216)
723	G550	Size	Dev and morph	Z-Dof	Small plant	724	(134-180)
725	G670	Size	Dev and morph	MYB-(R1)R2R3	Small plant	726	(14-122)
727	G760	Size	Dev and morph	NAC	Reduced size	728	(12-156)
729	G831	Size	Dev and morph	AKR	Reduced size	730	(470-591)
731	G864	Size	Dev and morph	AP2	Small plant	732	(119-186)
733	G884	Size	Dev and morph	WRKY	Reduced size	734	(227-285, 407-465)
735	G998	Size; seed oil and protein content	Dev and morph; seed biochemistry	RING/C3HC4	Reduced size; altered seed oil and protein content	736	(148-185)
737	G900	Size	Dev and morph	Z-CO-like	Reduced size	738	(6-28, 48-74)
739	G913	Size; flowering time	Dev and morph; flowering time	AP2	Small plant; late flowering	740	(62-128)
741	G937	Size	Dev and morph	GARP	Slightly reduced size	742	(197-246)
743	G960	Size	Dev and morph	NAC	Small plant	744	(13-156)
745	G991	Size; seed oil and protein content	Dev and morph; seed biochemistry	IAA	Slightly reduced size; altered seed oil and protein content	746	(7-14, 48-59, 82-115, 128-164)
747	G748	Stem; flowering time	Dev and morph; flowering time	Z-Dof	More vascular bundles in stem; late flowering	748	(112-140)
749	G247	Trichome; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Altered trichome distribution; altered seed protein content	750	(15-116)
751	G585	Trichome	Dev and morph	HLH/MYC	Reduced trichome density	752	(436-501)
753	G634	Trichome; seed protein content	Dev and morph; seed biochemistry	TH	Increased trichome density and size; altered seed protein content	754	(62-147, 189-245)
755	G676	Trichome	Dev and morph	MYB-(R1)R2R3	Reduced trichomes	756	(17-119)

Table 4

757	G682	Trichome	Dev and morph	MYB-related	Glabrous, lack of trichomes	758	(27-63)
759	G635	Variegation	Dev and morph	TH		760	(239-323)
761	G1068	Sugar sensing; seed oil and protein content	Sugar sensing; seed	AT-hook	Reduced cotyledon expansion in glucose	762	(143-150)
763	G1225	Sugar sensing; seed oil and protein content	biochemistry	HLH/MYC	Better germination on sucrose and glucose media; altered seed oil and protein content	764	(78-147)
765	G1337	Sugar sensing	Sugar sensing	Z-CO-like	Decreased germination on sucrose medium	766	(9-75)
767	G1759	Sugar sensing	Sugar sensing	MADS	Reduced germination on high glucose	768	(2-57)
769	G1804	Sugar sensing; flowering time	Sugar sensing; flowering time	bZIP	Altered sugar sensing; late flowering	770	(357-407)
771	G207	Sugar sensing	Sugar sensing	MYB-(R1)R2R3	Decreased germination on glucose medium	772	(6-106)
773	G218	Sugar sensing; seed oil content	Sugar sensing; seed	MYB-(R1)R2R3	Reduced cotyledon expansion in glucose; altered seed oil content	774	(TBD)
775	G241	Sugar sensing; seed oil and protein content	Sugar sensing; seed		Decreased germination and growth on glucose medium; decreased seed oil, altered protein content	776	(14-114)
777	G254	Sugar sensing	biochemistry	MYB-(R1)R2R3	Decreased germination and growth on glucose medium	778	(62-106)
779	G26	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	780	(67-134)
781	G263	Sugar sensing	Sugar sensing	HS	Decreased root growth on sucrose medium, root specific expression	782	(TBD)
783	G308	Sugar sensing	Sugar sensing	SCR	No germination on glucose medium	784	(270-274)
785	G38	Sugar sensing	Sugar sensing	AP2	Reduced germination on glucose medium	786	(76-143)
787	G43	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	788	(104-172)
789	G536	Sugar sensing	Sugar sensing	GF14	Decreased germination and growth on glucose medium	790	(226-233)
791	G567	Sugar sensing; seed oil and protein content	Sugar sensing; seed		Decreased seedling vigor on high glucose; altered seed oil and protein content	792	(210-270)
793	G680	Sugar sensing; flowering time	Sugar sensing; flowering time	bZIP	Reduced germination on glucose medium; late flowering	794	(24-70)
795	G667	Sugar sensing	Sugar sensing	MYB-related	Better seedling vigor on sucrose medium	796	(59-124)
797	G956	Sugar sensing	Sugar sensing	AP2	Reduced germination on glucose medium	798	(TBD)
799	G996	Sugar sensing	Sugar sensing	NAC	Reduced germination on glucose medium	800	(14-114)
		Seed glucosinolates, oil, protein content	Seed biochemistry	MYB-(R1)R2R3			
801	G1946	Seed oil content	Seed biochemistry	HS	Increase in M3950; increased oil content; decreased protein content	802	(32-130)
803	G217	Seed oil composition	Seed biochemistry	MYB-related	Increase in 20:2	804	(8-67)

Table 4

805	G2192	Seed oil composition	Seed biochemistry	bZIP-NIN	Altered composition	806	(600-700)
807	G504	Seed oil composition;	Seed biochemistry	NAC	Altered seed oil composition and content; altered seed protein content	808	(TBD)
809	G622	Seed oil composition	Seed biochemistry	ABI3/NIP-1	Decreased 18:2 fatty acid	810	(TBD)
811	G778	Seed oil composition	Seed biochemistry	HLH/MYC	Increased seed 18:1 fatty acid	812	(220-267)
813	G791	Seed oil composition	Seed biochemistry	HLH/MYC	Altered seed fatty acid composition	814	(75-143)
815	G861	Seed oil composition; seed oil content	Seed biochemistry	MADS	Increase in 16:1; altered seed oil content	816	(2-57)
817	G938	Seed oil composition	Seed biochemistry	EIL	Altered seed fatty acid composition	818	(96-104)
819	G965	Seed oil composition	Seed biochemistry	HB	Increase in 18:1	820	(423-486)
821	G1143	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	822	(33-82)
823	G1190	Seed oil content	Seed biochemistry	AKR	Increased content	824	(entire protein)
825	G1198	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	826	(173-223)
827	G1226	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	828	(115-174)
829	G1451	Seed oil content	Seed biochemistry	ARF	Altered seed oil content	830	(22-357)
831	G1478	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	Z-CO-like	Altered seed oil, protein content; late flowering	832	(32-76)
833	G1496	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	834	(184-248)
835	G1526	Seed oil content	Seed biochemistry	SWI/SNF	Increased seed oil content	836	(493-620, 864-1006)
837	G1543	Seed oil content	Seed biochemistry	HB	Decreased seed oil	838	(135-195)
839	G162	Seed oil and protein content	Seed biochemistry	MADS	Altered seed oil content; altered seed oil and protein content	840	(2-57)

Table 4

841	G1640	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Increased seed oil	842	(14-115)
843	G1644	Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil, protein content	844	(39-102)
845	G1646	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	846	(72-162)
847	G1672	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	848	(41-194)
849	G1677	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil, protein content	850	(17-181)
851	G1765	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	852	(20-140)
853	G1777	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Increased oil, decreased protein content	854	(124-247)
855	G1793	Seed oil content	Seed biochemistry	AP2	Increased seed oil content	856	(179-255, 281-349)
857	G180	Seed oil content	Seed biochemistry	WRKY	Decreased seed oil content	858	(118-174)
859	G192	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	WRKY	Altered seed oil and protein content; late flowering	860	(128-185)
861	G1948	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	862	(entire protein)
863	G2123	Seed oil and protein content	Seed biochemistry	GF14	Altered seed oil and protein content	864	(99-109)
865	G2138	Seed oil content	Seed biochemistry	AP2	Increased seed oil content	866	(TBD)
867	G2139	Seed oil content	Seed biochemistry	MADS	Increased seed content	868	(14-69)
869	G2343	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	870	(14-116)
871	G265	Seed oil and protein content	Seed biochemistry	HS	Altered seed oil and protein content	872	(11-105)
873	G2792	Seed oil content	Seed biochemistry	HLH/MYC	Increased seed oil content	874	(190-258)
875	G2830	Seed oil and protein content	Seed biochemistry	Z-C2H2	Altered seed oil and protein content	876	(245-266)
877	G286	Seed oil and protein content	Seed biochemistry	ENBP	Altered seed oil and protein content	878	(TBD)

Table 4

879	G291	Seed oil content	Seed biochemistry	MISC	Increased seed oil content	880	(132-160)
881	G427	Seed oil and protein content	Seed biochemistry	HB	Increased oil content; decreased protein content	882	(307-370)
883	G509	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	884	(13-169)
885	G519	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	886	(11-104)
887	G561	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	888	(248-308)
889	G590	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	890	(202-254)
891	G818	Seed oil content	Seed biochemistry	HS	Increased content	892	(70-162)
893	G849	Seed oil and protein content	Seed biochemistry	BPF-1	Increased seed oil, altered protein content	894	(324-413, 504-583)
895	G892	Seed oil and protein content	Seed biochemistry	RING/C3H2C3	Altered seed oil, protein content	896	(177-270)
897	G961	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	898	(15-140)
899	G1465	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	900	(242-306)
901	G425	Seed oil content	Seed biochemistry	HB	Altered seed oil content	902	(TBD)
903	G347	Seed oil and protein content	Seed biochemistry	Z-LSDlike	Altered seed oil and protein content	904	(9-39, 50-70, 80-127)
905	G1512	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	906	(39-93)
907	G2069	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	908	(TBD)
909	G1852	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	910	(1-601)
911	G1793	Seed oil content	Seed biochemistry	AP2	Altered seed oil content	912	(179-255, 281-349)
913	G761	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	914	(10-156)
915	G1056	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	916	(183-246)
917	G1447	Seed oil content	Seed biochemistry	MISC	Altered seed oil content	918	(3-54, 124-156)

Table 4

919	G323	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	920	(48-96)
921	G176	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	922	(117-173, 234-290)
923	G174	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	924	(111-166, 283-339)
925	G715	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	926	(60-132)
927	G588	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	928	(309-376)
929	G1758	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	930	(109-165)
931	G2148	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	932	(130-268)
933	G2379	Seed oil content	Seed biochemistry	TH	Altered seed oil content	934	(19-110, 173-232)
935	G1462	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	936	(TBD)
937	G1211	Seed oil and protein content	Seed biochemistry	MISC	Altered seed oil and protein content	938	(123-179)
939	G1048	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	940	(138-190)
941	G986	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	942	(146-203)
943	G789	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	944	(253-313)
945	G2085	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	946	(TBD)
947	G1783	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	948	(81-129)
949	G2072	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	950	(90-149)
951	G931	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil and protein content	952	(TBD)
953	G278	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	954	(2-593)
955	G2421	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	956	(9-110)
957	G2032	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	958	(entire protein)

Table 4

959	G1396	Seed oil and protein content	Seed biochemistry	S1FA	Altered seed oil and protein content	960	(TBD)
961	G619	Seed oil and protein content	Seed biochemistry	ARF	Altered seed oil and protein content	962	(64-406)
963	G2295	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	964	(2-57)
965	G312	Seed oil content	Seed biochemistry	SCR	Altered seed oil content	966	(320-336)
967	G1444	Seed oil and protein content	Seed biochemistry	GRF-like	Altered seed oil and protein content	968	(168-193)
969	G801	Seed oil content	Seed biochemistry	PCF	Altered seed oil content	970	(32-93)
971	G1950	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	972	(65-228)
973	G958	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	974	(7-156)
975	G1037	Seed oil and protein content	Seed biochemistry	GARP	Altered seed oil and protein content	976	(11-134, 200-248)
977	G2065	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	978	(TBD)
979	G2137	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	980	(109-168)
981	G746	Seed oil content	Seed biochemistry	RING/C3HC4	Altered seed oil content	982	(139-178)
983	G2701	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	984	(33-81, 129-183)
985	G1819	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	986	(46-188)
987	G1227	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	988	(183-244)
989	G2417	Seed oil content	Seed biochemistry	GARP	Altered seed oil content	990	(235-285)
991	G2116	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	992	(150-210)
993	G647	Seed oil content	Seed biochemistry	Z-C3H	Altered seed oil content	994	(77-192)
995	G974	Seed oil and protein content	Seed biochemistry	AP2	Altered seed oil and protein content	996	(81-140)
997	G1419	Seed protein content	Seed biochemistry	AP2	Increased seed protein	998	(69-137)

Table 4

999	G1634	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1000	(129-180)
1001	G1637	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1002	(109-173)
1003	G1818	Seed protein content; flowering time	Seed biochemistry; flowering time	CAAT	Increased protein content; late flowering	1004	(36-113)
1005	G1820	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil, protein content	1006	(70-133)
1007	G1903	Seed oil and protein content	Seed biochemistry	Z-Dof	Altered seed oil and protein content	1008	(134-180)
1009	G371	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	1010	(21-74)
1011	G597	Seed protein content	Seed biochemistry	AT-hook	Altered seed protein content	1012	(97-104, 137-144)
1013	G1009	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1014	(201-277, 303-371)
1015	G170	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1016	(2-57)
1017	G1768	Seed protein content	Seed biochemistry	SCR	Altered seed protein content	1018	(54-413)
1019	G185	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1020	(113-172)
1021	G1931	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1022	(114-170)
1023	G2543	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1024	(31-91)
1025	G264	Seed protein content	Seed biochemistry	HS	Altered seed protein content	1026	(24-114)
1027	G32	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1028	(17-84)
1029	G436	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1030	(22-85)
1031	G556	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1032	(83-143)
1033	G1420	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1034	(221-280)
1035	G1412	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1036	(17-159)

Table 4

1037	G738	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1038	(351-393)
1039	G2426	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1040	(14-114)
1041	G1524	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1042	(49-110)
1043	G1243	Seed protein content	Seed biochemistry	SWI/SNF	Altered seed protein content	1044	(216-609)
1045	G631	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1046	(TBD)
1047	G1909	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1048	(23-51)
1049	G1663	Seed protein content	Seed biochemistry	PCF	Altered seed protein content	1050	(TBD)
1051	G1231	Seed protein content	Seed biochemistry	Z-C4HC3	Altered seed protein content	1052	(TBD)
1053	G227	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1054	(13-112)
1055	G1842	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1056	(2-57)
1057	G1505	Seed protein content	Seed biochemistry	GATAZn	Altered seed protein content	1058	(TBD)
1059	G657	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1060	(TBD)
1061	G1959	Seed protein content	Seed biochemistry	GARP	Altered seed protein content	1062	(46-97)
1063	G2180	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1064	(7-156)
1065	G1817	Seed protein content	Seed biochemistry	PMR	Altered seed protein content	1066	(47-331)
1067	G1649	Seed protein content	Seed biochemistry	HLH/MYC	Altered seed protein content	1068	(225-295)
1069	G2131	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1070	(50-186, 112-183)
1071	G215	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1072	(TBD)
1073	G1508	Seed protein content	Seed biochemistry	GATAZn	Altered seed protein content	1074	(38-63)
1075	G2110	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1076	(239-298)

Table 4

1077	G2442	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1078	(220-246)
1079	G1051	Flowering time	Flowering time	bZIP	Late flowering	1080	(189-250)
1081	G1052	Flowering time	Flowering time	bZIP	Late flowering	1082	(201-261)
1083	G1079	Flowering time;	Flowering time;	BZIP2	Late flowering; altered seed protein content	1084	(1-50)
1085	G1335	Flowering time	Flowering time	Z-CLDSH	Late flowering, slow growth	1086	(24-43, 131-144, 185-203)
1087	G157	Flowering time	Flowering time	MADS	Altered flowering; significant overexpression delays	1088	(2-57)
1089	G1895	Flowering time	Flowering time	Z-Dof	Late flowering	1090	(55-110)
1091	G1900	Flowering time	Flowering time	Z-Dof	Late flowering	1092	(54-106)
1093	G2007	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Late flowering; altered seed protein content	1094	(TBD)
1095	G214	Flowering time	Flowering time	MYB-related	Late flowering	1096	(22-71)
1097	G2155	Flowering time	Flowering time	AT-hook	Late flowering	1098	(18-38)
1099	G234	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering, small plant	1100	(14-115)
1101	G361	Flowering time	Flowering time	Z-C2H2	Late flowering	1102	(43-63)
1103	G562	Flowering time	Flowering time	bZIP	Late flowering	1104	(253-315)
1105	G591	Flowering time	Flowering time	HLH/MYC	Late flowering	1106	(143-240)
1107	G8	Flowering time	Flowering time	AP2	Late flowering	1108	(151-217, 243-296)
1109	G859	Seed protein content	Seed biochemistry	MADS	Late flowering; altered seed protein content	1110	(TBD)
1111	G878	Flowering time	Flowering time	WRKY	Late flowering	1112	(250-305, 415-475)
1113	G971	Flowering time	Flowering time	AP2	Late flowering	1114	(120-186)
1115	G975	Flowering time; morphology: other	Flowering time; dev and morph	AP2	Late flowering; glossy leaves	1116	(4-71)
1117	G994	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering, small	1118	(14-123)
1119	G2347	Flowering time	Flowering time	SBP	Late flowering, small	1120	(60-136)
1121	G2010	Flowering time	Flowering time	SBP	Late flowering	1122	(53-127)

Table 5

SEQ ID NO	GID	Test Sequence ID	Smallest Sum Probability	Test Sequence Species	Test Sequence GenBank Annotation
859	G192	AW596933	7.70E-40	[Glycine max]	sj84f07.y1 Gm-c1034 Glycine max cDNA clone GENO
859	G192	AV423663	2.40E-39	[Lotus japonicus]	AV423663 Lotus japonicus young plants (two-
859	G192	BI422074	4.50E-34	[Lycopersicon esculentum]	EST532740 tomato callus, TAMU Lycop
859	G192	AW447931	1.40E-27	[Triticum aestivum]	BRY 1082 BRY Triticum aestivum cDNA clone
859	G192	BE998060	2.60E-24	[Medicago truncatula]	EST429783 GVSJN Medicago truncatula cDNA
859	G192	AC018727	1.70E-23	[Oryza sativa]	chromosome 10 clone OSJNBa0056G17, *** SEQUENC
859	G192	BG600477	1.00E-20	[Solanum tuberosum]	EST505372 cSTS Solanum tuberosum cDNA clo
859	G192	BG356878	2.80E-16	[Sorghum bicolor]	OV2 11 B04.g1 A002 Ovary 2 (OV2) Sorghum bi
859	G192	gi12039364	1.10E-31	[Oryza sativa]	putative DNA-binding protein.
859	G192	gi4894963	3.30E-14	[Avena sativa]	DNA-binding protein WRKY3.
859	G192	gi1432056	5.80E-14	[Petroselinum crispum]	WRKY3.
859	G192	gi4760596	2.60E-13	[Nicotiana tabacum]	DNA-binding protein NtWRKY3.
859	G192	gi11993901	1.40E-12	[Dactylis glomerata]	somatic embryogenesis related protein.
859	G192	gi927025	7.60E-09	[Cucumis sativus]	SPF1-like DNA-binding protein.
859	G192	gi13620227	8.40E-09	[Lycopersicon esculentum]	hypothetical protein.
859	G192	gi3420906	2.80E-08	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
859	G192	gi1159877	4.70E-08	[Avena fatua]	DNA-binding protein.
859	G192	gi484261	1.60E-07	[Ipomoea batatas]	SPF1 protein.
801	G1946	LPHSF8	1.10E-119	[Lycopersicon peruvianum]	L.peruvianum Lp-hsf8 mRNA for heat
801	G1946	AC087771	4.10E-112	[Medicago truncatula]	clone 8D15, *** SEQUENCING IN PROGRESS
801	G1946	LEHSF8	5.90E-103	[Lycopersicon esculentum]	L.esculentum Le-hsf8 gene for heat
801	G1946	AW569138	3.10E-75	[Glycine max]	si63g09.y1 Gm-r1030 Glycine max cDNA clone GENO
801	G1946	BG890899	1.30E-70	[Solanum tuberosum]	EST516750 cSTD Solanum tuberosum cDNA clo
801	G1946	AC027658	4.60E-53	[Oryza sativa]	subsp. japonica BAC nbxb0006113, chromosome 10
801	G1946	AV833112	4.90E-52	[Hordeum vulgare subsp. vulgare]	AV833112 K. Sato unpublished
801	G1946	gi19492	2.80E-121	[Lycopersicon peruvianum]	heat shock transcription factor 8
801	G1946	gi19260	5.10E-106	[Lycopersicon esculentum]	heat stress transcription factor
801	G1946	gi662924	2.00E-47	[Glycine max]	heat shock transcription factor 21.
801	G1946	gi5821138	9.70E-46	[Nicotiana tabacum]	heat shock factor.
801	G1946	gi11761077	2.90E-40	[Oryza sativa]	putative heat shock factor protein 1 (HSF 1)
801	G1946	gi886742	3.20E-40	[Zea mays]	heat shock factor.
801	G1946	gi7158882	2.70E-38	[Medicago sativa]	heat shock transcription factor.
801	G1946	gi3550588	1.90E-30	[Pisum sativum]	heat shock transcription factor (HSFA).

Table 5

801	G1946	gi100546	0.46	[Avena sativa]	avenin precursor - oat.
801	G1946	gi114190783	1	[Apium graveolens]	putative phloem transcription factor M1.
239	G375	AW696439	3.40E-33	[Medicago truncatula]	NF106B07ST1F1060 Developing stem Medica
239	G375	BG595870	1.90E-31	[Solanum tuberosum]	EST494548 cSTS Solanum tuberosum cDNA clo
239	G375	AI899263	3.70E-31	[Lycopersicon esculentum]	EST268706 tomato ovary, TAMU Lycope
239	G375	NTBBF3	4.00E-31	[Nicotiana tabacum]	N.tabacum mRNA for zinc finger protein, B
239	G375	BG405482	2.70E-30	[Glycine max]	sac44a11.y1 Gm-c1062 Glycine max cDNA clone GEN
239	G375	AB028130	3.30E-30	[Oryza sativa]	mRNA for Dof zinc finger protein, complete cds
239	G375	AB026297	7.30E-28	[Pisum sativum]	mRNA for elicitor-responsive Dof protein ERDP
239	G375	HVBPBF	1.10E-27	[Hordeum vulgare]	mRNA for DNA binding protein BPBF.
239	G375	BG263089	1.70E-27	[Triticum aestivum]	WHE2337_A02_A03ZS Wheat pre-anthesis spik
239	G375	ZMU82230	4.20E-27	[Zea mays]	endosperm-specific prolamin box binding factor (PB
239	G375	gi4996640	1.90E-37	[Oryza sativa]	Dof zinc finger protein.
239	G375	gi3777436	8.10E-35	[Hordeum vulgare]	DNA binding protein.
239	G375	gi2393775	1.10E-33	[Zea mays]	prolamin box binding factor.
239	G375	gi1360088	2.00E-33	[Nicotiana tabacum]	Zn finger protein.
239	G375	gi3790264	4.30E-32	[Triticum aestivum]	PBF protein.
239	G375	gi6092016	1.30E-29	[Pisum sativum]	elicitor-responsive Dof protein ERDP.
239	G375	gi7688355	5.60E-29	[Solanum tuberosum]	Dof zinc finger protein.
239	G375	gi1669341	4.60E-20	[Cucurbita maxima]	AOBP (ascorbate oxidase promoter-binding
239	G375	gi3929325	5.50E-18	[Dendrobium grex Madame Thong-In]	putative DNA-binding prot
239	G375	gi19547	5.50E-06	[Medicago sativa subsp. falcata]	environmental stress and a
273	G1255	AC087181	1.60E-46	[Oryza sativa]	chromosome 3 clone OSJNBa0018H01, *** SEQUENCI
273	G1255	BG239774	4.50E-33	[Glycine max]	sab74c03.y1 Gm-c1032 Glycine max cDNA clone GEN
273	G1255	BG321336	1.70E-32	[Descurainia sophia]	Ds01_06h10_A Ds01_AAFc_ECORC_cold_stress
273	G1255	AI772841	2.90E-30	[Lycopersicon esculentum]	EST253941 tomato resistant, Cornell
273	G1255	BF480245	4.60E-29	[Mesembryanthemum crystallinum]	L0-2152T3 Ice plant Lambda Un
273	G1255	AW688119	2.10E-28	[Medicago truncatula]	NF002E07ST1F1000 Developing stem Medica
273	G1255	BF266327	1.80E-26	[Hordeum vulgare]	HV_CEa0014N02f Hordeum vulgare seedling gre
273	G1255	AW671538	5.80E-25	[Sorghum bicolor]	LG1_348_B08.b1_A002 Light Grown 1 (LG1) Sor
273	G1255	BI072021	5.30E-20	[Populus tremula x Populus tremuloides]	C067P76U Populus stra
273	G1255	BG273908	4.90E-19	[Vitis vinifera]	EST 110 Green Grape berries Lambda Zap II Li
273	G1255	gi13702811	3.70E-52	[Oryza sativa]	putative zinc finger protein.
273	G1255	gi11037311	4.00E-21	[Brassica nigra]	constans-like protein.
273	G1255	gi2303683	1.10E-19	[Brassica napus]	unnamed protein product.
273	G1255	gi4091804	2.30E-18	[Malus x domestica]	CONSTANS-like protein 1.

Table 5

273	G1255	gi3341723	4.30E-18	[Raphanus sativus]	CONSTANS-like 1 protein.
273	G1255	gi10946337	5.20E-17	[Ipomoea nil]	CONSTANS-like protein.
273	G1255	gi4557093	3.30E-15	[Pinus radiata]	zinc finger protein.
273	G1255	gi8132543	0.97	[Chloroplast Zamia furfuracea]	cytochrome b559 alpha subunit
273	G1255	gi11795	0.99	[Nicotiana tabacum]	put. psbE protein (aa 1-83).
273	G1255	gi65646	0.99	[Chloroplast Nicotiana tabacum]	cytochrome b559 component p
557	G865	BE419451	3.70E-32	[Triticum aestivum]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
557	G865	AW560968	1.10E-28	[Medicago truncatula]	EST316016 DSIR Medicago truncatula cDNA
557	G865	AW782252	1.20E-26	[Glycine max]	sm03d11.y1 Gm-c1027 Glycine max cDNA clone GENO
557	G865	BI421895	3.60E-25	[Lycopersicon esculentum]	EST532561 tomato callus, TAMU Lycopersicon
557	G865	BE642320	1.60E-24	[Ceratopteris richardii]	Cri2_5_L17_SP6 Ceratopteris Spore Li
557	G865	BE494041	1.60E-24	[Secale cereale]	WHE1277_B09_D17ZS Secale cereale anther cDNA
557	G865	D39914	2.60E-24	[Oryza sativa]	RICS1576A Rice shoot Oryza sativa cDNA, mRNA s
557	G865	AV428124	9.00E-23	[Lotus japonicus]	AV428124 Lotus japonicus young plants (two-)
557	G865	TOBBY4D	1.80E-21	[Nicotiana tabacum]	Tobacco mRNA for EREBP-2, complete cds.
557	G865	gi1208495	2.40E-23	[Nicotiana tabacum]	ERF1.
557	G865	gi8809571	5.10E-23	[Nicotiana sylvestris]	ethylene-responsive element binding
557	G865	gi3342211	1.40E-22	[Lycopersicon esculentum]	Pti4.
557	G865	gi7528276	1.70E-22	[Mesembryanthemum crystallinum]	AP2-related transcription f
557	G865	gi15217291	7.80E-22	[Oryza sativa]	Putative AP2 domain containing protein.
557	G865	gi3264767	2.70E-21	[Prunus armeniaca]	AP2 domain containing protein.
557	G865	gi8980313	2.10E-20	[Catharanthus roseus]	AP2-domain DNA-binding protein.
557	G865	gi8571476	9.30E-20	[Atriplex hortensis]	apetala2 domain-containing protein.
557	G865	gi1688233	1.40E-19	[Solanum tuberosum]	DNA binding protein homolog.
557	G865	gi6478845	1.80E-19	[Matricaria chamomilla]	ethylene-responsive element binding
23	G2509	BH577856	2.50E-29	[Brassica oleracea]	BOHOJ67TR BOHO Brassica oleracea genomic
23	G2509	BM269574	5.90E-28	[Glycine max]	sak01e08.y1 Gm-c1074 Glycine max cDNA clone SOY
23	G2509	BE419451	2.20E-27	[Triticum aestivum]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
23	G2509	AI483636	7.80E-27	[Lycopersicon esculentum]	EST249507 tomato ovary, TAMU Lycopersicon
23	G2509	AW560968	8.90E-27	[Medicago truncatula]	EST316016 DSIR Medicago truncatula cDNA
23	G2509	BE642320	4.30E-26	[Ceratopteris richardii]	Cri2_5_L17_SP6 Ceratopteris Spore Li
23	G2509	AP003286	1.00E-25	[Oryza sativa]	chromosome 1 clone P0677H08, *** SEQUENCING IN
23	G2509	BE494041	3.20E-25	[Secale cereale]	WHE1277_B09_D17ZS Secale cereale anther cDNA
23	G2509	BE602106	1.10E-24	[Hordeum vulgare]	HVSMEH0102106f Hordeum vulgare 5-45 DAP spi
23	G2509	AV428124	1.00E-23	[Lotus japonicus]	AV428124 Lotus japonicus young plants (two-)
23	G2509	gi3264767	4.00E-27	[Prunus armeniaca]	AP2 domain containing protein.

Table 5

23	G2509	gi12003376	1.40E-23	[Nicotiana tabacum]	Avr/Cf-9 rapidly elicited protein 1.
23	G2509	gi14140141	2.30E-23	[Oryza sativa]	putative AP2-related transcription factor.
23	G2509	gi1688233	5.40E-23	[Solanum tuberosum]	DNA binding protein homolog.
23	G2509	gi4099921	2.60E-22	[Stylosanthes hamata]	EREBP-3 homolog.
23	G2509	gi8809571	7.80E-22	[Nicotiana sylvestris]	ethylene-responsive element binding
23	G2509	gi3342211	1.00E-21	[Lycopersicon esculentum]	Pti4.
23	G2509	gi7528276	2.70E-21	[Mesembryanthemum crystallinum]	AP2-related transcription f
23	G2509	gi17385636	1.90E-20	[Matricaria chamomilla]	ethylene-responsive element binding
23	G2509	gi18496063	3.30E-20	[Fagus sylvatica]	ethylene responsive element binding prote
1119	G2347	BI931517	5.30E-31	[Lycopersicon esculentum]	EST551406 tomato flower, 8 mm to pr
1119	G2347	BE058432	4.20E-29	[Glycine max]	sn16a06.y1 Gm-c1016 Glycine max cDNA clone GENO
1119	G2347	AMSPB1	1.80E-28	[Antirrhinum majus]	A.majus mRNA for squamosa-promoter bindin
1119	G2347	BG525285	5.70E-28	[Stevia rebaudiana]	48-3 Stevia field grown leaf cDNA Stevia
1119	G2347	L38193	4.60E-27	[Brassica rapa]	BNAF1025E Mustard flower buds Brassica rapa c
1119	G2347	BG455868	6.40E-27	[Medicago truncatula]	NF068F05PL1F1045 Phosphate starved leaf
1119	G2347	BG097153	1.70E-24	[Solanum tuberosum]	EST461672 potato leaves and petioles Sola
1119	G2347	BF482644	1.60E-23	[Triticum aestivum]	WHE2301-2304_A21_A21ZS Wheat pre-anthesis
1119	G2347	AW747167	2.30E-23	[Sorghum bicolor]	WS1_66_F11.b1_A002 Water-stressed 1 (WS1) S
1119	G2347	BG442540	2.50E-23	[Gossypium arboreum]	GA_Ea0017G06f Gossypium arboreum 7-10 d
1119	G2347	gi1183864	1.50E-31	[Antirrhinum majus]	squamosa-promoter binding protein 2.
1119	G2347	gi5931786	3.40E-25	[Zea mays]	SBP-domain protein 5.
1119	G2347	gi8468036	1.40E-21	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
1119	G2347	gi9087308	6.60E-09	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
1119	G2347	gi7209500	0.83	[Brassica rapa]	S-locus pollen protein.
43	G988	CRU303349	3.10E-208	[Capsella rubella]	ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 (pa
43	G988	A84072	4.50E-86	[Lycopersicon esculentum]	Sequence 1 from Patent WO9846759.
43	G988	A84080	3.30E-85	[Solanum tuberosum]	Sequence 9 from Patent WO9846759.
43	G988	AP003944	1.30E-57	[Oryza sativa]	chromosome 6 clone OJ1126 F05, *** SEQUENCING
43	G988	AX081276	2.80E-43	[Brassica napus]	Sequence 1 from Patent WO01093556.
43	G988	ZMA242530	1.50E-40	[Zea mays]	partial d8 gene for gibberellin response modulator
43	G988	AX005804	2.50E-37	[Triticum aestivum]	Sequence 13 from Patent WO9909174.
43	G988	AB048713	9.10E-33	[Pisum sativum]	PssCR mRNA for SCARECROW, complete cds.
43	G988	AW774515	2.00E-29	[Medicago truncatula]	EST333666 KV3 Medicago truncatula cDNA
43	G988	BE822458	1.20E-27	[Glycine max]	GM700017A20H12 Gm-r1070 Glycine max cDNA clone
43	G988	gi13620166	8.00E-211	[Capsella rubella]	hypothetical protein.
43	G988	gi4160441	1.40E-87	[Lycopersicon esculentum]	lateral suppressor protein.

Table 5

43	G988	gi10178637	2.20E-48	[Zea mays]	SCARECROW.
43	G988	gi6970472	1.20E-47	[Oryza sativa]	OsGAI.
43	G988	gi5640157	2.80E-45	[Triticum aestivum]	gibberellin response modulator.
43	G988	gi13170126	7.10E-45	[Brassica napus]	unnamed protein product.
43	G988	gi13365610	1.10E-40	[Pisum sativum]	SCARECROW.
43	G988	gi14318115	1.10E-14	[Zea mays subsp. mays]	gibberellin response modulator.
43	G988	gi14318165	7.30E-14	[Tripsacum dactyloides]	gibberellin response modulator.
43	G988	gi347457	2.40E-05	[Glycine max]	hydroxyproline-rich glycoprotein.
459	G2346	AMA011622	3.10E-35	[Antirrhinum majus]	mRNA for squamosa promoter binding
459	G2346	AW691786	1.80E-26	[Medicago truncatula]	NF044B06ST1F1000 Developing stem Medica
459	G2346	AQ273505	7.00E-25	[Oryza sativa]	nbxb003003f CUGI Rice BAC Library Oryza sativ
459	G2346	AW932595	7.90E-24	[Lycopersicon esculentum]	EST358438 tomato fruit mature green
459	G2346	BG593787	9.50E-24	[Solanum tuberosum]	EST492465 cSTS Solanum tuberosum cDNA clo
459	G2346	BG442540	1.00E-23	[Gossypium arboreum]	GA Ea0017G06f Gossypium arboreum 7-10 d
459	G2346	AZ919034	1.90E-23	[Zea mays]	1006013G02.x3 1006 - RescueMu Grid G Zea mays geno
459	G2346	BE596165	2.70E-23	[Sorghum bicolor]	PI1 50 D04.b1 A002 Pathogen induced 1 (PI1)
459	G2346	AI443033	2.30E-22	[Glycine max]	sa31a08.y1 Gm-c1004 Glycine max cDNA clone GENO
459	G2346	BF482644	4.30E-22	[Triticum aestivum]	WHE2301-2304 A21 A21ZS Wheat pre-anthesis
459	G2346	gi5931643	6.20E-45	[Antirrhinum majus]	squamosa promoter binding protein-homol
459	G2346	gi5931786	4.20E-26	[Zea mays]	SBP-domain protein 5.
459	G2346	gi8468036	3.30E-14	[Zea sativa]	Similar to Arabidopsis thaliana chromosome 2
459	G2346	gi9087308	8.30E-08	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
285	G1354	BG128374	2.90E-58	[Lycopersicon esculentum]	EST474020 tomato shoot/meristem Lyc
285	G1354	BE202831	1.90E-56	[Medicago truncatula]	EST402853 KV1 Medicago truncatula cDNA
285	G1354	AI161918	6.60E-55	[Populus tremula x Populus tremuloides]	A009P50U Hybrid aspen
285	G1354	AB028186	1.20E-53	[Oryza sativa]	mRNA for OsNAC7 protein, complete cds.
285	G1354	BE060921	8.00E-50	[Hordeum vulgare]	HVSMEG0013N15f Hordeum vulgare pre-anthesis
285	G1354	AF402603	1.50E-42	[Phaseolus vulgaris]	NAC domain protein NAC2 mRNA, complete c
285	G1354	BE357920	1.60E-42	[Sorghum bicolor]	DG1 23 F03.b1 A002 Dark Grown 1 (DG1) Sorgh
285	G1354	PHRNANAM	3.60E-42	[Petunia x hybrida]	P.hybrida mRNA encoding NAM protein.
285	G1354	AW185617	5.30E-40	[Glycine max]	se80b05.y1 Gm-c1023 Glycine max cDNA clone GENO
285	G1354	gi6006373	4.50E-63	[Oryza sativa]	Similar to NAM like protein (AC005310).
285	G1354	gi15148914	2.30E-44	[Phaseolus vulgaris]	NAC domain protein NAC2.
285	G1354	gi14485513	3.50E-44	[Solanum tuberosum]	putative NAC domain protein.
285	G1354	gi1279640	5.90E-44	[Petunia x hybrida]	NAM.
285	G1354	gi6175246	5.20E-41	[Lycopersicon esculentum]	jasmonic acid 2.

Table 5

285	G1354	gi4218535	5.10E-39	[Triticum sp.]	GRAB1 protein.
285	G1354	gi6732158	5.10E-39	[Triticum monococcum]	unnamed protein product.
285	G1354	gi7716952	3.30E-35	[Medicago truncatula]	NAC1.
285	G1354	gi4996349	2.50E-26	[Nicotiana tabacum]	NAC-domain protein.
285	G1354	gi2982275	3.10E-14	[Picea mariana]	ATAF1-like protein.
119	G1063	BH700922	4.50E-90	[Brassica oleracea]	BOMMZ07TR BO 2 3 KB Brassica oleracea gen
119	G1063	BE451174	2.40E-41	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
119	G1063	AW832545	2.00E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
119	G1063	AP004693	5.90E-37	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
119	G1063	AP004462	4.40E-32	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
119	G1063	AT002234	8.90E-32	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
119	G1063	BF263465	5.40E-25	[Hordeum vulgare]	HV CEa0006N02f Hordeum vulgare seedling gre
119	G1063	BG557011	4.20E-22	[Sorghum bicolor]	EM1 41 E02.g1 A002 Embryo 1 (EM1) Sorghum b
119	G1063	BG842856	3.10E-21	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
119	G1063	BG559930	1.40E-18	[Sorghum propinquum]	RHIZ2 75 D09.g1 A003 Rhizome2 (RHIZ2) So
119	G1063	gi15528743	4.20E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
119	G1063	gi6166283	8.10E-10	[Pinus taeda]	helix-loop-helix protein 1A.
119	G1063	gi11045087	8.80E-09	[Brassica napus]	putative protein.
119	G1063	gi10998404	7.10E-08	[Petunia x hybrida]	anthocyanin 1.
119	G1063	gi99441	2.60E-07	[Volvox carteri]	sulfated surface glycoprotein 185 - Volvox
119	G1063	gi1142621	5.00E-07	[Phaseolus vulgaris]	phaseolin G-box binding protein PG2.
119	G1063	gi166428	8.10E-07	[Antirrhinum majus]	DEL.
119	G1063	gi1247386	9.50E-07	[Nicotiana glauca]	PRP2.
119	G1063	gi82091	1.00E-06	[Lycopersicon esculentum]	hydroxyproline-rich glycoprotein
119	G1063	gi1486263	1.40E-06	[Catharanthus roseus]	extensin.
129	G2143	BH650724	3.00E-88	[Brassica oleracea]	BOMIW43TR BO 2 3 KB Brassica oleracea gen
129	G2143	AW832545	1.50E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
129	G2143	BE451174	3.50E-40	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
129	G2143	AP004693	4.00E-38	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
129	G2143	AP004584	6.30E-33	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
129	G2143	AT002234	3.00E-31	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
129	G2143	BF263465	2.90E-26	[Hordeum vulgare]	HV CEa0006N02f Hordeum vulgare seedling gre
129	G2143	BG557011	2.60E-22	[Sorghum bicolor]	EM1 41 E02.g1 A002 Embryo 1 (EM1) Sorghum b
129	G2143	BG842856	3.50E-20	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
129	G2143	BG559930	6.10E-18	[Sorghum propinquum]	RHIZ2 75 D09.g1 A003 Rhizome2 (RHIZ2) So
129	G2143	gi15528743	5.50E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.

Table 5

129	G2143	gi1086538	7.60E-09	[Oryza rufipogon]	transcriptional activator Rb homolog.
129	G2143	gi6166283	1.10E-08	[Pinus taeda]	helix-loop-helix protein 1A.
129	G2143	gi1142621	4.60E-07	[Phaseolus vulgaris]	phasolin G-box binding protein PG2.
129	G2143	gi3399777	5.20E-07	[Glycine max]	symbiotic ammonium transporter; nodulin.
129	G2143	gi5923912	6.10E-07	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
129	G2143	gi10998404	9.20E-07	[Petunia x hybrida]	anthocyanin 1.
129	G2143	gi4321762	5.20E-06	[Zea mays]	transcription factor MYC7E.
129	G2143	gi166428	6.00E-06	[Antirrhinum majus]	DEL.
129	G2143	gi527665	7.40E-06	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	BH511840	6.70E-62	[Brassica oleracea]	BOGRJ19TR BOGR Brassica oleracea genomic
133	G2557	BE347811	3.70E-46	[Glycine max]	sp05h10.y1 Gm-c1041 Glycine max cDNA clone GENO
133	G2557	AP003141	2.40E-33	[Oryza sativa]	genomic DNA, chromosome 1, PAC clone:P0002B05,
133	G2557	BF263465	3.00E-31	[Hordeum vulgare]	HV_CEA0006N02f Hordeum vulgare seedling gre
133	G2557	AT002234	6.60E-27	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
133	G2557	BG557011	6.40E-26	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
133	G2557	AP004462	7.90E-26	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clob
133	G2557	BE451174	3.90E-25	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
133	G2557	BG842856	5.60E-22	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
133	G2557	BG559930	7.00E-14	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
133	G2557	gi15289790	2.40E-36	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
133	G2557	gi3399777	2.60E-06	[Glycine max]	symbiotic ammonium transporter; nodulin.
133	G2557	gi4206118	1.10E-05	[Mesembryanthemum crystallinum]	transporter homolog.
133	G2557	gi6166283	1.30E-05	[Pinus taeda]	helix-loop-helix protein 1A.
133	G2557	gi527655	3.70E-05	[Pennisetum glaucum]	myc-like regulatory R gene product.
133	G2557	gi5923912	3.70E-05	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
133	G2557	gi527661	7.80E-05	[Phyllostachys acuta]	myc-like regulatory R gene product.
133	G2557	gi527665	9.50E-05	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	gi1086538	0.0001	[Oryza rufipogon]	transcriptional activator Rb homolog.
133	G2557	gi5669656	0.00013	[Lycopersicon esculentum]	ER333 protein.
697	G2430	BF632520	1.90E-14	[Medicago truncatula]	NF039A08DT1F1054 Drought Medicago trunc
697	G2430	AW396912	1.20E-13	[Glycine max]	sg64g09.y1 Gm-c1007 Glycine max cDNA clone GENO
697	G2430	D41804	4.50E-13	[Oryza sativa]	RICS4626A Rice shoot Oryza sativa cDNA, mRNA s
697	G2430	BE214029	2.60E-10	[Hordeum vulgare]	HV_CEB0001P06f Hordeum vulgare seedling gre
697	G2430	AW564570	2.70E-10	[Sorghum bicolor]	LG1_296_E01.b1_A002 Light Grown 1 (LG1) Sor
697	G2430	BG129795	5.40E-10	[Lycopersicon esculentum]	EST475441 tomato shoot/meristem Lyc
697	G2430	AB060130	5.40E-09	[Zea mays]	ZmRR8 mRNA for response regulator 8, complete cds.

Table 5

697	G2430	BF587105	2.50E-05	[Sorghum propinquum]	FM1_32_C05.b1_A003 Floral-Induced Merist
697	G2430	A1163121	0.3	[Populus tremula x Populus tremuloides]	A033P70U Hybrid aspen
697	G2430	BG595628	0.46	[Solanum tuberosum]	EST494306 cSTS Solanum tuberosum cDNA clo
697	G2430	gi13661174	5.40E-18	[Zea mays]	response regulator 8.
697	G2430	gi15289981	0.028	[Oryza sativa]	hypothetical protein.
697	G2430	gi6942190	0.12	[Mesembryanthemum crystallinum]	CDPK substrate protein 1; C
697	G2430	gi4519671	0.2	[Nicotiana tabacum]	transfactor.
831	G1478	BF275913	1.50E-20	[Gossypium arboreum]	GA_Eb0025C07f Gossypium arboreum 7-10 d
831	G1478	BG157399	6.50E-19	[Glycine max]	sab36g12.y1 Gm-c1026 Glycine max cDNA clone GEN
831	G1478	C95300	2.20E-10	[Citrus unshiu]	C95300 Citrus unshiu Miyagawa-wase maturation
831	G1478	AW034552	2.70E-10	[Lycopersicon esculentum]	EST278168 tomato callus, TAMU Lycop
831	G1478	BI070429	3.40E-10	[Populus tremula x Populus tremuloides]	C037P68U Populus tra
831	G1478	AF016011	5.10E-09	[Brassica napus]	CONSTANS homolog (Bn9CON10) gene, complete c
831	G1478	BE598912	6.20E-09	[Sorghum bicolor]	PI1_84_H11.b1_A002 Pathogen induced 1 (PI1)
831	G1478	BG605313	6.80E-09	[Triticum aestivum]	WHE2331_C04_F07ZS Wheat pre-anthesis spk
831	G1478	BE558327	8.90E-09	[Hordeum vulgare]	HV_CEB0017D19f Hordeum vulgare seedling gre
831	G1478	BG647091	1.20E-08	[Medicago truncatula]	EST508710 HOGA Medicago truncatula cDNA
831	G1478	gi2895188	4.70E-11	[Brassica napus]	CONSTANS homolog.
831	G1478	gi3618308	1.50E-09	[Oryza sativa]	zinc finger protein.
831	G1478	gi11037308	4.70E-09	[Brassica nigra]	constans-like protein.
831	G1478	gi3341723	1.30E-08	[Raphanus sativus]	CONSTANS-like 1 protein.
831	G1478	gi4091806	1.50E-07	[Malus x domestica]	CONSTANS-like protein 2.
831	G1478	gi10946337	3.10E-07	[Ipomoea nil]	CONSTANS-like protein.
831	G1478	gi4557093	1.40E-05	[Pinus radiata]	zinc finger protein.
831	G1478	gi619312	0.9	[Capparis masaiikai]	mabinlin III B-chain=sweet protein mabi
831	G1478	gi4732091	1	[Zea mays]	bundle sheath defective protein 2.
831	G1478	gi4699629	1	[Nicotiana glauca]	Chain A, Putative Ancestral Protein Encod
579	G681	BG128147	6.80E-41	[Lycopersicon esculentum]	EST473793 tomato shoot/meristem Lyc
579	G681	BF054497	1.50E-39	[Solanum tuberosum]	EST439727 potato leaves and petioles Sola
579	G681	BE054276	8.40E-39	[Gossypium arboreum]	GA_Ea0002O18f Gossypium arboreum 7-10 d
579	G681	BG269414	4.00E-38	[Mesembryanthemum crystallinum]	L0-3478T3 Ice plant Lambda Un
579	G681	BF620286	7.40E-38	[Hordeum vulgare]	HVSMIEc0019F08f Hordeum vulgare seedling sho
579	G681	BE490032	1.00E-37	[Triticum aestivum]	WHE0364_C04_E08ZS Wheat cold-stressed see
579	G681	BI542536	1.40E-36	[Zea mays]	949021A03.y1 949 - Juvenile leaf and shoot cDNA fr
579	G681	BF425254	7.20E-36	[Glycine max]	su42c10.y1 Gm-c1068 Glycine max cDNA clone GENO
579	G681	AW672062	3.20E-34	[Sorghum bicolor]	LG1_354_G05.b1_A002 Light Grown 1 (LG1) Sor

Table 5

579	G681	BG448527	1.00E-33	[Medicago truncatula]	NF036F04RT1F1032 Developing root Medica
579	G681	gi13346188	9.10E-37	[Gossypium hirsutum]	GhMYB25.
579	G681	gi20563	6.30E-36	[Petunia x hybrida]	protein 1.
579	G681	gi485867	1.20E-34	[Antirrhinum majus]	mixta.
579	G681	gi2605617	1.70E-32	[Oryza sativa]	OSMYB1.
579	G681	gi1430846	2.00E-31	[Lycopersicon esculentum]	myb-related transcription factor.
579	G681	gi6651292	2.20E-30	[Pimpinella brachycarpa]	myb-related transcription factor.
579	G681	gi15042116	4.90E-30	[Zea mays subsp. parviglumis]	Cl protein.
579	G681	gi82730	6.10E-30	[Zea mays]	transforming protein (myb) homolog (clone Zm38)
579	G681	gi5139806	8.30E-30	[Glycine max]	GmMYB29A2.
579	G681	gi19055	1.10E-29	[Hordeum vulgare]	MybHv5.
611	G878	AF096299	6.20E-90	[Nicotiana tabacum]	DNA-binding protein 2 (WRKY2) mRNA, compl
611	G878	CUSSLD8	1.80E-83	[Cucumis sativus]	SPF1-like DNA-binding protein mRNA, complet
611	G878	AF193802	3.50E-63	[Oryza sativa]	zinc finger transcription factor WRKY1 mRNA, c
611	G878	AX192162	2.20E-62	[Glycine max]	Sequence 9 from Patent WO0149840.
611	G878	IPBSPF1P	3.80E-58	[Ipomoea batatas]	Sweet potato mRNA for SPF1 protein, complet
611	G878	AFABF1	2.00E-56	[Avena fatua]	A. fatua mRNA for DNA-binding protein (clone ABF
611	G878	LES303343	7.20E-55	[Lycopersicon esculentum]	mRNA for hypothetical protein (ORF
611	G878	AX192164	4.00E-54	[Triticum aestivum]	Sequence 11 from Patent WO0149840.
611	G878	AF080595	2.10E-53	[Pimpinella brachycarpa]	zinc finger protein (ZFP1) mRNA, com
611	G878	PCU48831	2.30E-53	[Petroselinum crispum]	DNA-binding protein WRKY1 mRNA, comple
611	G878	gi4322940	3.30E-128	[Nicotiana tabacum]	DNA-binding protein 2.
611	G878	gi927025	1.10E-109	[Cucumis sativus]	SPF1-like DNA-binding protein.
611	G878	gi6689916	1.50E-74	[Oryza sativa]	zinc finger transcription factor WRKY1.
611	G878	gi484261	1.10E-66	[Ipomoea batatas]	SPF1 protein.
611	G878	gi1159877	2.30E-63	[Avena fatua]	DNA-binding protein.
611	G878	gi13620227	4.60E-63	[Lycopersicon esculentum]	hypothetical protein.
611	G878	gi5917653	1.70E-56	[Petroselinum crispum]	zinc-finger type transcription facto
611	G878	gi4894965	5.00E-56	[Avena sativa]	DNA-binding protein WRKY1.
611	G878	gi3420906	8.70E-56	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
611	G878	gi13620168	4.20E-22	[Capsella rubella]	hypothetical protein.
47	G374	AP004457	1.20E-73	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
47	G374	AP004693	1.90E-73	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
47	G374	BH552835	1.30E-62	[Brassica oleracea]	BOHGT561TR BOHG Brassica oleracea genomic
47	G374	BG128229	6.50E-55	[Lycopersicon esculentum]	EST473875 tomato shoot/meristem Lyc
47	G374	BG646959	3.20E-46	[Medicago truncatula]	EST508578 HOGA Medicago truncatula cDNA

Table 5

47	G374	BG890162	8.70E-41	[Solanum tuberosum]	EST516013 cSTD Solanum tuberosum cDNA clo
47	G374	AW179366	6.00E-38	[Zea mays]	618046G06.y1 618 - Inbred Tassel cDNA Library Zea
47	G374	BF473206	1.50E-32	[Triticum aestivum]	WHE0922 G12 M24ZS Wheat 5-15 DAP spike cD
47	G374	AW761011	2.90E-29	[Glycine max]	sl61g11.y1 Gm-c1027 Glycine max cDNA clone GENO
47	G374	AJ436050	1.50E-27	[Hordeum vulgare]	AJ436050 S00007 Hordeum vulgare cDNA clone
47	G374	gi422012	0.8	[Sorghum bicolor]	lipid transfer protein - sorghum (fragmen
47	G374	gi1827893	1	[Zea mays]	Maize Nonspecific Lipid Transfer Protein Complex

Traits of interest

Examples of some of the traits that may be desirable in plants, and that may be provided by transforming the plants with the presently disclosed sequences, are listed in Table 6.

Table 6. Genes, traits and utilities that affect plant characteristics

<u>Trait Category</u>	<u>Traits</u>	Transcription factor genes that <u>impact traits</u>	<u>Utility</u> <u>Gene effect on:</u>
Resistance and tolerance	Salt stress resistance	G22; G196; G226; G303; G312; G325; G353; G482; G545; G801; G867; G884; G922; G926; G1452; G1794; G1820; G1836; G1843; G1863; G2053; G2110; G2140; G2153; G2379; G2701; G2713; G2719; G2789	Germination rate, survivability, yield; extended growth range
	Osmotic stress resistance	G47; G175; G188; G303; G325; G353; G489; G502; G526; G921; G922; G926; G1069; G1089; G1452; G1794; G1930; G2140; G2153; G2379; G2701; G2719; G2789;	Germination rate, survivability, yield
	Cold stress resistance; cold germination	G256; G394; G664; G864; G1322; G2130	Germination, growth, earlier planting
	Tolerance to freezing	G303; G325; G353; G720; G912; G913; G1794; G2053; G2140; G2153; G2379; G2701; G2719; G2789	Survivability, yield, appearance, extended range
	Heat stress resistance	G3; G464; G682; G864; G964;	Germination,

		G1305; G1645; G2130 G2430	growth, later planting
	Drought, low humidity resistance	G303; G325; G353; G720; G912; G926; G1452; G1794; G1820; G1843; G2053; G2140; G2153; G2379; G2583; G2701; G2719; G2789	Survivability, yield, extended range
	Radiation resistance	G1052	Survivability, vigor, appearance
	Decreased herbicide sensitivity	G343; G2133; G2517	Resistant to increased herbicide use
	Increased herbicide sensitivity	G374; G877; G1519	Use as a herbicide target
	Oxidative stress	G477; G789; G1807; G2133; G2517	Improved yield, appearance, reduced senescence
	Light response	G183; G354; G375; G1062; G1322; G1331; G1488; G1494; G1521; G1786; G1794; G2144; G2555;	Germination, growth, development, flowering time
Development, morphology	Overall plant architecture	G24; G27; G31; G33; G47; G147; G156; G160; G182; G187; G195; G196; G211; G221; G237; G280; G342; G352; G357; G358; G360; G362; G364; G365; G367; G373; G377; G396; G431; G447; G479; G546; G546; G551; G578; G580; G596; G615; G617; G620; G625;	Vascular tissues, lignin content; cell wall content; appearance

		G638; G658; G716; G725; G727; G730; G740; G770; G858; G865; G869; G872; G904; G910; G912; G920; G939; G963; G977; G979; G987; G988; G993; G1007; G1010; G1014; G1035; G1046; G1049; G1062; G1069; G1070; G1076; G1089; G1093; G1127; G1131; G1145; G1229; G1246; G1304; G1318; G1320; G1330; G1331; G1352; G1354; G1360; G1364; G1379; G1384; G1399; G1415; G1417; G1442; G1453; G1454; G1459; G1460; G1471; G1475; G1477; G1487; G1487; G1492; G1499; G1499; G1531; G1540; G1543; G1543; G1544; G1548; G1584; G1587; G1588; G1589; G1636; G1642; G1747; G1749; G1749; G1751; G1752; G1763; G1766; G1767; G1778; G1789; G1790; G1791; G1793; G1794; G1795; G1800; G1806; G1811; G1835; G1836; G1838; G1839; G1843; G1853; G1855; G1865; G1881; G1882; G1883; G1884; G1891; G1896; G1898; G1902; G1904; G1906; G1913; G1914; G1925; G1929; G1930; G1954; G1958; G1965; G1976; G2057; G2107; G2133; G2134; G2151; G2154; G2157; G2181;	
--	--	--	--

		G2290; G2299; G2340; G2340; G2346; G2373; G2376; G2424; G2465; G2505; G2509; G2512; G2513; G2519; G2520; G2533; G2534; G2573; G2589; G2687; G2720; G2787; G2789; G2893	
	Size: increased stature	G189; G1073; G1435; G2430	
	Size: reduced stature or dwarfism	G3; G5; G21; G23; G39; G165; G184; G194; G258; G280; G340; G343; G353; G354; G362; G363; G370; G385; G396; G439; G440; G447; G450; G550; G557; G599; G636; G652; G670; G671; G674; G729; G760; G804; G831; G864; G884; G898; G900; G912; G913; G922; G932; G937; G939; G960; G962; G977; G991; G1000; G1008; G1020; G1023; G1053; G1067; G1075; G1137; G1181; G1198; G1228; G1266; G1267; G1275; G1277; G1309; G1311; G1314; G1317; G1322; G1323; G1326; G1332; G1334; G1367; G1381; G1382; G1386; G1421; G1488; G1494; G1537; G1545; G1560; G1586; G1641; G1652; G1655; G1671; G1750; G1756; G1757; G1782; G1786; G1794; G1839; G1845; G1879; G1886; G1888; G1933; G1939; G1943; G1944; G2011; G2094; G2115;	Ornamental; small stature provides wind resistance; creation of dwarf varieties

		G2130; G2132; G2144; G2145; G2147; G2156; G2294; G2313; G2344; G2431; G2510; G2517; G2521; G2893; G2893	
	Fruit size and number	G362	Biomass, yield, cotton boll fiber density
	Flower structure, inflorescence	G47; G259; G353; G354; G671; G732; G988; G1000; G1063; G1140; G1326; G1449; G1543; G1560; G1587; G1645; G1947; G2108; G2143; G2893	Ornamental horticulture; production of saffron or other edible flowers
	Number and development of trichomes	G225; G226; G247; G362; G585; G634; G676; G682; G1014; G1332; G1452; G1795; G2105	Resistance to pests and desiccation; essential oil production
	Seed size, color, and number	G156; G450; G584; G652; G668; G858; G979; G1040; G1062; G1145; G1255; G1494; G1531; G1534; G1594; G2105; G2114;	Yield
	Root development, modifications	G9; G1482; G1534; G1794; G1852; G2053; G2136; G2140	
	Modifications to root hairs	G225; G226	Nutrient, water uptake, pathogen resistance
	Apical dominance	G559; G732; G1255; G1275; G1411; G1488; G1635; G2452; G2509	Ornamental horticulture
	Branching patterns	G568; G988; G1548	Ornamental horticulture, knot reduction, improved

			windscreen
	Leaf shape, color, modifications	G375; G377; G428; G438; G447; G464; G557; G577; G599; G635; G671; G674; G736; G804; G903; G977; G921; G922; G1038; G1063; G1067; G1073; G1075; G1146; G1152; G1198; G1267; G1269; G1452; G1484; G1586; G1594; G1767; G1786; G1792; G1886; G2059; G2094; G2105; G2113; G2117; G2143; G2144; G2431; G2452; G2465; G2587; G2583; G2724;	Appealing shape or shiny leaves for ornamental agriculture, increased biomass or photosynthesis
	Silique	G1134	Ornamental
	Stem morphology	G47; G438; G671; G748; G988; G1000	Ornamental; digestibility
	Shoot modifications	G390; G391	Ornamental stem bifurcations
Disease, Pathogen Resistance	Bacterial	G211; G347; G367; G418; G525; G545; G578; G1049	Yield, appearance, survivability, extended range
	Fungal	G19; G28; G28; G28; G147; G188; G207; G211; G237; G248; G278; G347; G367; G371; G378; G409; G477; G545; G545; G558; G569; G578; G591; G594; G616; G789; G805; G812; G865; G869; G872; G881; G896; G940; G1047; G1049; G1064; G1084; G1196; G1255; G1266;	Yield, appearance, survivability, extended range

		G1363; G1514; G1756; G1792; G1792; G1792; G1792; G1880; G1919; G1919; G1927; G1927; G1936; G1936; G1950; G2069; G2130; G2380; G2380; G2555	
Nutrients	Increased tolerance to nitrogen-limited soils	G225; G226; G1792	
	Increased tolerance to phosphate-limited soils	G419; G545; G561; G1946	
	Increased tolerance to potassium-limited soils	G561; G911	
Hormonal	Hormone sensitivity	G12; G546; G926; G760; G913; G926; G1062; G1069; G1095; G1134; G1330; G1452; G1666; G1820; G2140; G2789	Seed dormancy, drought tolerance; plant form, fruit ripening
Seed biochemistry	Production of seed prenyl lipids, including tocopherol	G214; G259; G490; G652; G748; G883; G1052; G1328; G1930; G2509; G2520	Antioxidant activity, vitamin E
	Production of seed sterols	G20	Precursors for human steroid hormones; cholesterol modulators
	Production of seed glucosinolates	G353; G484; G674; G1272; G1506; G1897; G1946; G2113; G2117; G2155; G2290; G2340	Defense against insects; putative anticancer activity; undesirable in

			animal feeds
	Modified seed oil content	G162; G162; G180; G192; G241; G265; G286; G291; G427; G509; G519; G561; G567; G590; G818; G849; G892; G961; G974; G1063; G1143; G1190; G1198; G1226; G1229; G1323; G1451; G1471; G1478; G1496; G1526; G1543; G1640; G1644; G1646; G1672; G1677; G1750; G1765; G1777; G1793; G1838; G1902; G1946; G1948; G2059; G2123; G2138; G2139; G2343; G2792; G2830	Vegetable oil production; increased caloric value for animal feeds; lutein content
	Modified seed oil composition	G217; G504; G622; G778; G791; G861; G869; G938; G965; G1417; G2192	Heat stability, digestibility of seed oils
	Modified seed protein content	G162; G226; G241; G371; G427; G509; G567; G597; G732; G849; G865; G892; G963; G988; G1323; G1323; G1419; G1478; G1488; G1634; G1637; G1641; G1644; G1652; G1677; G1777; G1777; G1818; G1820; G1903; G1909; G1946; G1946; G1958; G2059; G2117; G2417; G2509	Reduced caloric value for humans
Leaf biochemistry	Production of flavonoids	G1666*	Ornamental pigment production; pathogen resistance; health

			benefits
	Production of leaf glucosinolates	G264; G353; G484; G652; G674; G681; G1069; G1198; G1322; G1421; G1657; G1794; G1897; G1946; G2115; G2117; G2144; G2155; G2155; G2340; G2512; G2520; G2552	Defense against insects; putative anticancer activity; undesirable in animal feeds
	Production of diterpenes	G229	Induction of enzymes involved in alkaloid biosynthesis
	Production of anthocyanin	G546	Ornamental pigment
	Production of leaf phytosterols, inc. stigmastanol, campesterol	G561; G2131; G2424	Precursors for human steroid hormones; cholesterol modulators
	Leaf fatty acid composition	G214; G377; G861; G962; G975; G987; G1266; G1337; G1399; G1465; G1512; G2136; G2147; G2192	Nutritional value; increase in waxes for disease resistance
	Production of leaf prenyl lipids, including tocopherol	G214; G259; G280; G652; G987; G1543; G2509; G2520	Antioxidant activity, vitamin E
Biochemistry, general	Production of miscellaneous secondary metabolites	G229; G663	
	Sugar, starch, hemicellulose composition,	G158; G211; G211; G237; G242; G274; G598; G1012; G1266; G1309; G1309; G1641; G1765; G1865; G2094; G2094;	Food digestibility, hemicellulose & pectin content; fiber content; plant

		G2589; G2589	tensile strength, wood quality, pathogen resistance, pulp production; tuber starch content
Sugar sensing	Plant response to sugars	G26; G38; G43; G207; G218; G241; G254; G263; G308; G536; G567; G567; G680; G867; G912; G956; G996; G1068; G1225; G1314; G1314; G1337; G1759; G1804; G2153; G2379	Photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, senescence
Growth, Reproduction	Plant growth rate and development	G447; G617; G674; G730; G917; G937; G1035; G1046; G1131; G1425; G1452; G1459; G1492; G1589; G1652; G1879; G1943; G2430; G2431; G2465; G2521	Faster growth, increased biomass or yield, improved appearance; delay in bolting
	Embryo development	G167	
	Seed germination rate	G979; G1792; G2130	Yield
	Plant, seedling vigor	G561; G2346	Survivability, yield
	Senescence; cell death	G571; G636; G878; G1050; G1463; G1749; G1944; G2130; G2155; G2340; G2383	Yield, appearance; response to pathogens;
	Modified fertility	G39; G340; G439; G470; G559; G615; G652; G671; G779; G962; G977; G988; G1000; G1063; G1067; G1075;	Prevents or minimizes escape of the pollen of GMOs

		G1266; G1311; G1321; G1326; G1367; G1386; G1421; G1453; G1471; G1453; G1560; G1594; G1635; G1750; G1947; G2011; G2094; G2113; G2115; G2130; G2143; G2147; G2294; G2510; G2893	
	Early flowering	G147; G157; G180; G183; G183; G184; G185; G208; G227; G294; G390; G390; G390; G391; G391; G427; G427; G490; G565; G590; G592; G720; G789; G865; G898; G898; G989; G989; G1037; G1037; G1142; G1225; G1225; G1226; G1242; G1305; G1305; G1380; G1380; G1480; G1480; G1488; G1494; G1545; G1545; G1649; G1706; G1760; G1767; G1767; G1820; G1841; G1841; G1842; G1843; G1843; G1946; G1946; G2010; G2030; G2030; G2144; G2144; G2295; G2295; G2347; G2348; G2348; G2373; G2373; G2509; G2509; G2555; G2555	Faster generation time; synchrony of flowering; potential for introducing new traits to single variety
	Delayed flowering	G8; G47; G192; G214; G234; G361; G362; G562; G568; G571; G591; G680; G736; G748; G859; G878; G910; G912; G913; G971; G994; G1051; G1052; G1073; G1079; G1335; G1435; G1452; G1478;	Delayed time to pollen production of GMO plants; synchrony of flowering; increased yield

		G1789; G1804; G1865; G1865; G1895; G1900; G2007; G2133; G2155; G2291; G2465	
	Extended flowering phase	G1947	
	Flower and leaf development	G259; G353; G377; G580; G638 G652; G858; G869; G917; G922; G932; G1063; G1075; G1140; G1425; G1452; G1499; G1548; G1645; G1865; G1897; G1933; G2094; G2124; G2140; G2143; G2535; G2557	Ornamental applications; decreased fertility
	Flower abscission	G1897	Ornamental: longer retention of flowers

* When co-expressed with G669 and G663

Significance of modified plant traits

Currently, the existence of a series of maturity groups for different latitudes represents a major barrier to the introduction of new valuable traits. Any trait (e.g. disease resistance) has to be bred into each of the different maturity groups separately, a laborious and costly exercise. The availability of single strain, which could be grown at any latitude, would therefore greatly increase the potential for introducing new traits to crop species such as soybean and cotton.

For many of the traits, listed in Table 6 and below, that may be conferred to plants, a single transcription factor gene may be used to increase or decrease, advance or delay, or improve or prove deleterious to a given trait. For example, overexpression of a transcription factor gene that naturally occurs in a plant may cause early flowering relative to non-transformed or wild-type plants. By knocking out the gene, or suppressing the gene (with, for example, antisense suppression) the plant may experience delayed flowering. Similarly, overexpressing or suppressing one or more genes can impart significant differences in production of plant products,

such as different fatty acid ratios. Thus, suppressing a gene that causes a plant to be more sensitive to cold may improve a plant's tolerance of cold.

Salt stress resistance. Soil salinity is one of the more important variables that determines where a plant may thrive. Salinity is especially important for the successful cultivation of crop plants, particular in many parts of the world that have naturally high soil salt concentrations, or where the soil has been over-utilized. Thus, presently disclosed transcription factor genes that provide increased salt tolerance during germination, the seedling stage, and throughout a plant's life cycle would find particular value for imparting survivability and yield in areas where a particular crop would not normally prosper.

Osmotic stress resistance. Presently disclosed transcription factor genes that confer resistance to osmotic stress may increase germination rate under adverse conditions, which could impact survivability and yield of seeds and plants.

Cold stress resistance. The potential utility of presently disclosed transcription factor genes that increase tolerance to cold is to confer better germination and growth in cold conditions. The germination of many crops is very sensitive to cold temperatures. Genes that would allow germination and seedling vigor in the cold would have highly significant utility in allowing seeds to be planted earlier in the season with a high rate of survivability. Transcription factor genes that confer better survivability in cooler climates allow a grower to move up planting time in the spring and extend the growing season further into autumn for higher crop yields.

Tolerance to freezing. The presently disclosed transcription factor genes that impart tolerance to freezing conditions are useful for enhancing the survivability and appearance of plants conditions or conditions that would otherwise cause extensive cellular damage. Thus, germination of seeds and survival may take place at temperatures significantly below that of the mean temperature required for germination of seeds and survival of non-transformed plants. As with salt tolerance, this has the added benefit of increasing the potential range of a crop plant into regions in which it would otherwise succumb. Cold tolerant transformed plants may also be

planted earlier in the spring or later in autumn, with greater success than with non-transformed plants.

Heat stress tolerance. The germination of many crops is also sensitive to high temperatures. Presently disclosed transcription factor genes that provide increased heat tolerance are generally useful in producing plants that germinate and grow in hot conditions, may find particular use for crops that are planted late in the season, or extend the range of a plant by allowing growth in relatively hot climates.

Drought, low humidity tolerance. Strategies that allow plants to survive in low water conditions may include, for example, reduced surface area or surface oil or wax production. A number of presently disclosed transcription factor genes increase a plant's tolerance to low water conditions and provide the benefits of improved survivability, increased yield and an extended geographic and temporal planting range.

Radiation resistance. Presently disclosed transcription factor genes have been shown to increase lutein production. Lutein, like other xanthophylls such as zeaxanthin and violaxanthin, are important in the protection of plants against the damaging effects of excessive light. Lutein contributes, directly or indirectly, to the rapid rise of non-photochemical quenching in plants exposed to high light. Increased tolerance of field plants to visible and ultraviolet light impacts survivability and vigor, particularly for recent transplants. Also affected are the yield and appearance of harvested plants or plant parts. Crop plants engineered with presently disclosed transcription factor genes that cause the plant to produce higher levels of lutein therefore would have improved photoprotection, leading to less oxidative damage and increase vigor, survivability and higher yields under high light and ultraviolet light conditions.

Decreased herbicide sensitivity. Presently disclosed transcription factor genes that confer resistance or tolerance to herbicides (e.g., glyphosate) may find use in providing means to increase herbicide applications without detriment to desirable plants. This would allow for the increased use of a particular herbicide in a local

environment, with the effect of increased detriment to undesirable species and less harm to transgenic, desirable cultivars.

Increased herbicide sensitivity. Knockouts of a number of the presently disclosed transcription factor genes have been shown to be lethal to developing embryos. Thus, these genes are potentially useful as herbicide targets.

Oxidative stress. In plants, as in all living things, abiotic and biotic stresses induce the formation of oxygen radicals, including superoxide and peroxide radicals. This has the effect of accelerating senescence, particularly in leaves, with the resulting loss of yield and adverse effect on appearance. Generally, plants that have the highest level of defense mechanisms, such as, for example, polyunsaturated moieties of membrane lipids, are most likely to thrive under conditions that introduce oxidative stress (e.g., high light, ozone, water deficit, particularly in combination). Introduction of the presently disclosed transcription factor genes that increase the level of oxidative stress defense mechanisms would provide beneficial effects on the yield and appearance of plants. One specific oxidizing agent, ozone, has been shown to cause significant foliar injury, which impacts yield and appearance of crop and ornamental plants. In addition to reduced foliar injury that would be found in ozone resistant plant created by transforming plants with some of the presently disclosed transcription factor genes, the latter have also been shown to have increased chlorophyll fluorescence (Yu-Sen Chang et al. Bot. Bull. Acad. Sin. (2001) 42: 265-272).

Heavy metal tolerance. Heavy metals such as lead, mercury, arsenic, chromium and others may have a significant adverse impact on plant respiration. Plants that have been transformed with presently disclosed transcription factor genes that confer improved resistance to heavy metals, through, for example, sequestering or reduced uptake of the metals will show improved vigor and yield in soils with relatively high concentrations of these elements. Conversely, transgenic transcription factors may also be introduced into plants to confer an increase in heavy metal uptake, which may benefit efforts to clean up contaminated soils.

Light response. Presently disclosed transcription factor genes that modify a plant's response to light may be useful for modifying a plant's growth or

development, for example, photomorphogenesis in poor light, or accelerating flowering time in response to various light intensities, quality or duration to which a non-transformed plant would not similarly respond. Examples of such responses that have been demonstrated include leaf number and arrangement, and early flower bud appearances.

Overall plant architecture. Several presently disclosed transcription factor genes have been introduced into plants to alter numerous aspects of the plant's morphology. For example, it has been demonstrated that a number of transcription factors may be used to manipulate branching, such as the means to modify lateral branching, a possible application in the forestry industry. Transgenic plants have also been produced that have altered cell wall content, lignin production, flower organ number, or overall shape of the plants. Presently disclosed transcription factor genes transformed into plants may be used to affect plant morphology by increasing or decreasing internode distance, both of which may be advantageous under different circumstances. For example, for fast growth of woody plants to provide more biomass, or fewer knots, increased internode distances are generally desirable. For improved wind screening of shrubs or trees, or harvesting characteristics of, for example, members of the Gramineae family, decreased internode distance may be advantageous. These modifications would also prove useful in the ornamental horticulture industry for the creation of unique phenotypic characteristics of ornamental plants.

Increased stature. For some ornamental plants, the ability to provide larger varieties may be highly desirable. For many plants, including fruit-bearing trees or trees and shrubs that serve as view or wind screens, increased stature provides obvious benefits. Crop species may also produce higher yields on larger cultivars.

Reduced stature or dwarfism. Presently disclosed transcription factor genes that decrease plant stature can be used to produce plants that are more resistant to damage by wind and rain, or more resistant to heat or low humidity or water deficit. Dwarf plants are also of significant interest to the ornamental horticulture industry, and particularly for home garden applications for which space availability may be limited.

Fruit size and number. Introduction of presently disclosed transcription factor genes that affect fruit size will have desirable impacts on fruit size and number, which may comprise increases in yield for fruit crops, or reduced fruit yield, such as when vegetative growth is preferred (e.g., with bushy ornamentals, or where fruit is undesirable, as with ornamental olive trees).

Flower structure, inflorescence, and development. Presently disclosed transgenic transcription factors have been used to create plants with larger flowers or arrangements of flowers that are distinct from wild-type or non-transformed cultivars. This would likely have the most value for the ornamental horticulture industry, where larger flowers or interesting presentations generally are preferred and command the highest prices. Flower structure may have advantageous effects on fertility, and could be used, for example, to decrease fertility by the absence, reduction or screening of reproductive components. One interesting application for manipulation of flower structure, for example, by introduced transcription factors could be in the increased production of edible flowers or flower parts, including saffron, which is derived from the stigmas of *Crocus sativus*.

Number and development of trichomes. Several presently disclosed transcription factor genes have been used to modify trichome number and amount of trichome products in plants. Trichome glands on the surface of many higher plants produce and secrete exudates that give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or ant-microbial or they may act as allergens or irritants to protect against herbivores. Trichomes have also been suggested to decrease transpiration by decreasing leaf surface air flow, and by exuding chemicals that protect the leaf from the sun.

Seed size, color and number. The introduction of presently disclosed transcription factor genes into plants that alter the size or number of seeds may have a significant impact on yield, both when the product is the seed itself, or when biomass of the vegetative portion of the plant is increased by reducing seed production. In the case of fruit products, it is often advantageous to modify a plant to have reduced size

or number of seeds relative to non-transformed plants to provide seedless or varieties with reduced numbers or smaller seeds. Presently disclosed transcription factor genes have also been shown to affect seed size, including the development of larger seeds. Seed size, in addition to seed coat integrity, thickness and permeability, seed water content and by a number of other components including antioxidants and oligosaccharides, may affect seed longevity in storage. This would be an important utility when the seed of a plant is the harvested crops, as with, for example, peas, beans, nuts, etc. Presently disclosed transcription factor genes have also been used to modify seed color, which could provide added appeal to a seed product.

Root development, modifications. By modifying the structure or development of roots by transforming into a plant one or more of the presently disclosed transcription factor genes, plants may be produced that have the capacity to thrive in otherwise unproductive soils. For example, grape roots that extend further into rocky soils, or that remain viable in waterlogged soils, would increase the effective planting range of the crop. It may be advantageous to manipulate a plant to produce short roots, as when a soil in which the plant will be growing is occasionally flooded, or when pathogenic fungi or disease-causing nematodes are prevalent.

Modifications to root hairs. Presently disclosed transcription factor genes that increase root hair length or number potentially could be used to increase root growth or vigor, which might in turn allow better plant growth under adverse conditions such as limited nutrient or water availability.

Apical dominance. The modified expression of presently disclosed transcription factors that control apical dominance could be used in ornamental horticulture, for example, to modify plant architecture.

Branching patterns. Several presently disclosed transcription factor genes have been used to manipulate branching, which could provide benefits in the forestry industry. For example, reduction in the formation of lateral branches could reduce knot formation. Conversely, increasing the number of lateral branches could provide utility when a plant is used as a windscreen, or may also provide ornamental advantages.

Leaf shape, color and modifications. It has been demonstrated in laboratory experiments that overexpression of some of the presently disclosed transcription factors produced marked effects on leaf development. At early stages of growth, these transgenic seedlings developed narrow, upward pointing leaves with long petioles, possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements. Other transcription factor genes can be used to increase plant biomass; large size would be useful in crops where the vegetative portion of the plant is the marketable portion.

Siliques. Genes that later silique conformation in brassicates may be used to modify fruit ripening processes in brassicates and other plants, which may positively affect seed or fruit quality.

Stem morphology and shoot modifications. Laboratory studies have demonstrated that introducing several of the presently disclosed transcription factor genes into plants can cause stem bifurcations in shoots, in which the shoot meristems split to form two or three separate shoots. This unique appearance would be desirable in ornamental applications.

Diseases, pathogens and pests. A number of the presently disclosed transcription factor genes have been shown to or are likely to confer resistance to various plant diseases, pathogens and pests. The offending organisms include fungal pathogens *Fusarium oxysporum*, *Botrytis cinerea*, *Sclerotinia sclerotiorum*, and *Erysiphe orontii*. Bacterial pathogens to which resistance may be conferred include *Pseudomonas syringae*. Other problem organisms may potentially include nematodes, mollicutes, parasites, or herbivorous arthropods. In each case, one or more transformed transcription factor genes may provide some benefit to the plant to help prevent or overcome infestation. The mechanisms by which the transcription factors work could include increasing surface waxes or oils, surface thickness, local senescence, or the activation of signal transduction pathways that regulate plant defense in response to attacks by herbivorous pests (including, for example, protease inhibitors).

Increased tolerance of plants to nutrient-limited soils. Presently disclosed transcription factor genes introduced into plants may provide the means to improve uptake of essential nutrients, including nitrogenous compounds, phosphates, potassium, and trace minerals. The effect of these modifications is to increase the seedling germination and range of ornamental and crop plants. The utilities of presently disclosed transcription factor genes conferring tolerance to conditions of low nutrients also include cost savings to the grower by reducing the amounts of fertilizer needed, environmental benefits of reduced fertilizer runoff; and improved yield and stress tolerance. In addition, this gene could be used to alter seed protein amounts and/or composition that could impact yield as well as the nutritional value and production of various food products.

Hormone sensitivity. One or more of the presently disclosed transcription factor genes have been shown to affect plant abscisic acid (ABA) sensitivity. This plant hormone is likely the most important hormone in mediating the adaptation of a plant to stress. For example, ABA mediates conversion of apical meristems into dormant buds. In response to increasingly cold conditions, the newly developing leaves growing above the meristem become converted into stiff bud scales that closely wrap the meristem and protect it from mechanical damage during winter. ABA in the bud also enforces dormancy; during premature warm spells, the buds are inhibited from sprouting. Bud dormancy is eliminated after either a prolonged cold period of cold or a significant number of lengthening days. Thus, by affecting ABA sensitivity, introduced transcription factor genes may affect cold sensitivity and survivability. ABA is also important in protecting plants from drought tolerance.

Several other of the present transcription factor genes have been used to manipulate ethylene signal transduction and response pathways. These genes can thus be used to manipulate the processes influenced by ethylene, such as seed germination or fruit ripening, and to improve seed or fruit quality.

Production of seed and leaf prenyl lipids, including tocopherol. Prenyl lipids play a role in anchoring proteins in membranes or membranous organelles. Thus modifying the prenyl lipid content of seeds and leaves could affect membrane integrity and function. A number of presently disclosed transcription factor genes

have been shown to modify the tocopherol composition of plants. Tocopherols have both anti-oxidant and vitamin E activity.

Production of seed and leaf phytosterols: Presently disclosed transcription factor genes that modify levels of phytosterols in plants may have at least two utilities. First, phytosterols are an important source of precursors for the manufacture of human steroid hormones. Thus, regulation of transcription factor expression or activity could lead to elevated levels of important human steroid precursors for steroid semi-synthesis. For example, transcription factors that cause elevated levels of campesterol in leaves, or sitosterols and stigmasterols in seed crops, would be useful for this purpose. Phytosterols and their hydrogenated derivatives phytostanols also have proven cholesterol-lowering properties, and transcription factor genes that modify the expression of these compounds in plants would thus provide health benefits.

Production of seed and leaf glucosinolates. Some glucosinolates have anti-cancer activity; thus, increasing the levels or composition of these compounds by introducing several of the presently disclosed transcription factors might be of interest from a nutraceutical standpoint. (3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters might be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Modified seed oil content. The composition of seeds, particularly with respect to seed oil amounts and/or composition, is very important for the nutritional value and production of various food and feed products. Several of the presently disclosed transcription factor genes in seed lipid saturation that alter seed oil content could be used to improve the heat stability of oils or to improve the nutritional quality of seed oil, by, for example, reducing the number of calories in seed, increasing the number of calories in animal feeds, or altering the ratio of saturated to unsaturated lipids comprising the oils.

Seed and leaf fatty acid composition. A number of the presently disclosed transcription factor genes have been shown to alter the fatty acid composition in plants, and seeds in particular. This modification may find particular value for improving the nutritional value of, for example, seeds or whole plants. Dietary fatty acids ratios have been shown to have an effect on, for example, bone integrity and remodeling (see, for example, Weiler, H.A., *Pediatr Res* (2000) 47:5 692-697). The ratio of dietary fatty acids may alter the precursor pools of long-chain polyunsaturated fatty acids that serve as precursors for prostaglandin synthesis. In mammalian connective tissue, prostaglandins serve as important signals regulating the balance between resorption and formation in bone and cartilage. Thus dietary fatty acid ratios altered in seeds may affect the etiology and outcome of bone loss.

Modified seed protein content. As with seed oils, the composition of seeds, particularly with respect to protein amounts and/or composition, is very important for the nutritional value and production of various food and feed products. A number of the presently disclosed transcription factor genes modify the protein concentrations in seeds would provide nutritional benefits, and may be used to prolong storage, increase seed pest or disease resistance, or modify germination rates.

Production of flavonoids in leaves and other plant parts. Expression of presently disclosed transcription factor genes that increase flavonoid production in plants, including anthocyanins and condensed tannins, may be used to alter in pigment production for horticultural purposes, and possibly increasing stress resistance. Flavonoids have antimicrobial activity and could be used to engineer pathogen resistance. Several flavonoid compounds have health promoting effects such as the inhibition of tumor growth and cancer, prevention of bone loss and the prevention of the oxidation of lipids. Increasing levels of condensed tannins, whose biosynthetic pathway is shared with anthocyanin biosynthesis, in forage legumes is an important agronomic trait because they prevent pasture bloat by collapsing protein foams within the rumen. For a review on the utilities of flavonoids and their derivatives, refer to Dixon et al. (1999) *Trends Plant Sci.* 4:394-400.

Production of diterpenes in leaves and other plant parts. Depending on the plant species, varying amounts of diverse secondary biochemicals (often lipophilic

terpenes) are produced and exuded or volatilized by trichomes. These exotic secondary biochemicals, which are relatively easy to extract because they are on the surface of the leaf, have been widely used in such products as flavors and aromas, drugs, pesticides and cosmetics. Thus, the overexpression of genes that are used to produce diterpenes in plants may be accomplished by introducing transcription factor genes that induce said overexpression. One class of secondary metabolites, the diterpenes, can effect several biological systems such as tumor progression, prostaglandin synthesis and tissue inflammation. In addition, diterpenes can act as insect pheromones, termite allomones, and can exhibit neurotoxic, cytotoxic and antimitotic activities. As a result of this functional diversity, diterpenes have been the target of research several pharmaceutical ventures. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity.

Production of anthocyanin in leaves and other plant parts. Several presently disclosed transcription factor genes can be used to alter anthocyanin production in numerous plant species. The potential utilities of these genes include alterations in pigment production for horticultural purposes, and possibly increasing stress resistance in combination with another transcription factor.

Production of miscellaneous secondary metabolites. Microarray data suggests that flux through the aromatic amino acid biosynthetic pathways and primary and secondary metabolite biosynthetic pathways are up-regulated. Presently disclosed transcription factors have been shown to be involved in regulating alkaloid biosynthesis, in part by up-regulating the enzymes indole-3-glycerol phosphatase and strictosidine synthase. Phenylalanine ammonia lyase, chalcone synthase and trans-cinnamate mono-oxygenase are also induced, and are involved in phenylpropanoid biosynthesis.

Sugar, starch, hemicellulose composition. Overexpression of the presently disclosed transcription factors that affect sugar content resulted in plants with altered leaf insoluble sugar content. Transcription factors that alter plant cell wall composition have several potential applications including altering food digestibility, plant tensile strength, wood quality, pathogen resistance and in pulp production. The

potential utilities of a gene involved in glucose-specific sugar sensing are to alter energy balance, photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, and senescence.

Hemicellulose is not desirable in paper pulps because of its lack of strength compared with cellulose. Thus modulating the amounts of cellulose vs. hemicellulose in the plant cell wall is desirable for the paper/lumber industry. Increasing the insoluble carbohydrate content in various fruits, vegetables, and other edible consumer products will result in enhanced fiber content. Increased fiber content would not only provide health benefits in food products, but might also increase digestibility of forage crops. In addition, the hemicellulose and pectin content of fruits and berries affects the quality of jam and catsup made from them. Changes in hemicellulose and pectin content could result in a superior consumer product.

Plant response to sugars and sugar composition. In addition to their important role as an energy source and structural component of the plant cell, sugars are central regulatory molecules that control several aspects of plant physiology, metabolism and development. It is thought that this control is achieved by regulating gene expression and, in higher plants, sugars have been shown to repress or activate plant genes involved in many essential processes such as photosynthesis, glyoxylate metabolism, respiration, starch and sucrose synthesis and degradation, pathogen response, wounding response, cell cycle regulation, pigmentation, flowering and senescence. The mechanisms by which sugars control gene expression are not understood.

Because sugars are important signaling molecules, the ability to control either the concentration of a signaling sugar or how the plant perceives or responds to a signaling sugar could be used to control plant development, physiology or metabolism. For example, the flux of sucrose (a disaccharide sugar used for systemically transporting carbon and energy in most plants) has been shown to affect gene expression and alter storage compound accumulation in seeds. Manipulation of the sucrose signaling pathway in seeds may therefore cause seeds to have more protein, oil or carbohydrate, depending on the type of manipulation. Similarly, in tubers, sucrose is converted to starch which is used as an energy store. It is thought that sugar signaling pathways may partially determine the levels of starch synthesized

in the tubers. The manipulation of sugar signaling in tubers could lead to tubers with a higher starch content.

Thus, the presently disclosed transcription factor genes that manipulate the sugar signal transduction pathway may lead to altered gene expression to produce plants with desirable traits. In particular, manipulation of sugar signal transduction pathways could be used to alter source-sink relationships in seeds, tubers, roots and other storage organs leading to increase in yield.

Plant growth rate and development. A number of the presently disclosed transcription factor genes have been shown to have significant effects on plant growth rate and development. These observations have included, for example, more rapid or delayed growth and development of reproductive organs. This would provide utility for regions with short or long growing seasons, respectively. Accelerating plant growth would also improve early yield or increase biomass at an earlier stage, when such is desirable (for example, in producing forestry products).

Embryo development. Presently disclosed transcription factor genes that alter embryo development has been used to alter seed protein and oil amounts and/or composition which is very important for the nutritional value and production of various food products. Seed shape and seed coat may also be altered by these genes, which may provide for improved storage stability.

Seed germination rate. A number of the presently disclosed transcription factor genes have been shown to modify seed germination rate, including when the seeds are in conditions normally unfavorable for germination (e.g., cold, heat or salt stress, or in the presence of ABA), and may thus be used to modify and improve germination rates under adverse conditions.

Plant, seedling vigor. Seedlings transformed with presently disclosed transcription factors have been shown to possess larger cotyledons and appeared somewhat more advanced than control plants. This indicates that the seedlings developed more rapidly than the control plants. Rapid seedling development is likely to reduce loss due to diseases particularly prevalent at the seedling stage (e.g.,

damping off) and is thus important for survivability of plants germinating in the field or in controlled environments.

Senescence, cell death. Presently disclosed transcription factor genes may be used to alter senescence responses in plants. Although leaf senescence is thought to be an evolutionary adaptation to recycle nutrients, the ability to control senescence in an agricultural setting has significant value. For example, a delay in leaf senescence in some maize hybrids is associated with a significant increase in yields and a delay of a few days in the senescence of soybean plants can have a large impact on yield. Delayed flower senescence may also generate plants that retain their blossoms longer and this may be of potential interest to the ornamental horticulture industry.

Modified fertility. Plants that overexpress a number of the presently disclosed transcription factor genes have been shown to possess reduced fertility. This could be a desirable trait, as it could be exploited to prevent or minimize the escape of the pollen of genetically modified organisms (GMOs) into the environment.

Early and delayed flowering. Presently disclosed transcription factor genes that accelerate flowering could have valuable applications in such programs since they allow much faster generation times. In a number of species, for example, broccoli, cauliflower, where the reproductive parts of the plants constitute the crop and the vegetative tissues are discarded, it would be advantageous to accelerate time to flowering. Accelerating flowering could shorten crop and tree breeding programs. Additionally, in some instances, a faster generation time might allow additional harvests of a crop to be made within a given growing season. A number of *Arabidopsis* genes have already been shown to accelerate flowering when constitutively expressed. These include LEAFY, APETALA1 and CONSTANS (Mandel, M. et al., 1995, Nature 377, 522-524; Weigel, D. and Nilsson, O., 1995, Nature 377, 495-500; Simon et al., 1996, Nature 384, 59-62).

By regulating the expression of potential flowering using inducible promoters, flowering could be triggered by application of an inducer chemical. This would allow flowering to be synchronized across a crop and facilitate more efficient harvesting. Such inducible systems could also be used to tune the flowering of crop varieties to

different latitudes. At present, species such as soybean and cotton are available as a series of maturity groups that are suitable for different latitudes on the basis of their flowering time (which is governed by day-length). A system in which flowering could be chemically controlled would allow a single high-yielding northern maturity group to be grown at any latitude. In southern regions such plants could be grown for longer, thereby increasing yields, before flowering was induced. In more northern areas, the induction would be used to ensure that the crop flowers prior to the first winter frosts.

In a sizeable number of species, for example, root crops, where the vegetative parts of the plants constitute the crop and the reproductive tissues are discarded, it would be advantageous to delay or prevent flowering. Extending vegetative development with presently disclosed transcription factor genes could thus bring about large increases in yields.. Prevention of flowering might help maximize vegetative yields and prevent escape of genetically modified organism (GMO) pollen.

Extended flowering phase. Presently disclosed transcription factors that extend flowering time have utility in engineering plants with longer-lasting flowers for the horticulture industry, and for extending the time in which the plant is fertile.

Flower and leaf development. Presently disclosed transcription factor genes have been used to modify the development of flowers and leaves. This could be advantageous in the development of new ornamental cultivars that present unique configurations. In addition, some of these genes have been shown to reduce a plant's fertility, which is also useful for helping to prevent development of pollen of GMOs.

Flower abscission. Presently disclosed transcription factor genes introduced into plants have been used to retain flowers for longer periods. This would provide a significant benefit to the ornamental industry, for both cut flowers and woody plant varieties (of, for example, maize), as well as have the potential to lengthen the fertile period of a plant, which could positively impact yield and breeding programs.

A listing of specific effects and utilities that the presently disclosed transcription factor genes have on plants, as determined by direct observation and assay analysis, is provided in Table 4.

XVI. Antisense and Co-suppression

In addition to expression of the nucleic acids of the invention as gene replacement or plant phenotype modification nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, e.g., as a further mechanism for modulating plant phenotype. That is, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997) Antisense Technology: A Practical Approach IRL Press at Oxford University Press, Oxford, U.K.. In general, sense or anti-sense sequences are introduced into a cell, where they are optionally amplified, e.g., by transcription. Such sequences include both simple oligonucleotide sequences and catalytic sequences such as ribozymes.

For example, a reduction or elimination of expression (i.e., a “knock-out”) of a transcription factor or transcription factor homologue polypeptide in a transgenic plant, e.g., to modify a plant trait, can be obtained by introducing an antisense construct corresponding to the polypeptide of interest as a cDNA. For antisense suppression, the transcription factor or homologue cDNA is arranged in reverse orientation (with respect to the coding sequence) relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length cDNA or gene, and need not be identical to the cDNA or gene found in the plant type to be transformed. Typically, the antisense sequence need only be capable of hybridizing to the target gene or RNA of interest. Thus, where the introduced sequence is of shorter length, a higher degree of homology to the endogenous transcription factor sequence will be needed for effective antisense suppression. While antisense sequences of various lengths can be utilized, preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides.

Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous transcription factor gene in the plant cell.

Suppression of endogenous transcription factor gene expression can also be achieved using a ribozyme. Ribozymes are RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 and U.S. Patent No. 5,543,508. Synthetic ribozyme sequences including antisense RNAs can be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that hybridize to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Suppression of endogenous transcription factor gene expression can also be achieved using RNA interference, or RNAi. RNAi is a post-transcriptional, targeted gene-silencing technique that uses double-stranded RNA (dsRNA) to incite degradation of messenger RNA (mRNA) containing the same sequence as the dsRNA (Constans, (2002) *The Scientist* 16:36). Small interfering RNAs, or siRNAs are produced in at least two steps: an endogenous ribonuclease cleaves longer dsRNA into shorter, 21-23 nucleotide-long RNAs. The siRNA segments then mediate the degradation of the target mRNA (Zamore, (2001) *Nature Struct. Biol.*, 8:746-50). RNAi has been used for gene function determination in a manner similar to antisense oligonucleotides (Constans, (2002) *The Scientist* 16:36). Expression vectors that continually express siRNAs in transiently and stably transfected have been engineered to express small hairpin RNAs (shRNAs), which get processed in vivo into siRNA-like molecules capable of carrying out gene-specific silencing (Brummelkamp et al., (2002) *Science* 296:550-553, and Paddison, et al. (2002) *Genes & Dev.* 16:948-958). Post-transcriptional gene silencing by double-stranded RNA is discussed in further detail by Hammond et al. (2001) *Nature Rev Gen* 2: 110-119, Fire et al. (1998) *Nature* 391: 806-811 and Timmons and Fire (1998) *Nature* 395: 854.

Vectors in which RNA encoded by a transcription factor or transcription factor homologue cDNA is over-expressed can also be used to obtain co-suppression of a corresponding endogenous gene, e.g., in the manner described in U.S. Patent No.

5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire transcription factor cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous transcription factor gene of interest. However, as with antisense suppression, the suppressive efficiency will be enhanced as specificity of hybridization is increased, e.g., as the introduced sequence is lengthened, and/or as the sequence similarity between the introduced sequence and the endogenous transcription factor gene is increased.

Vectors expressing an untranslatable form of the transcription factor mRNA, e.g., sequences comprising one or more stop codon, or nonsense mutation) can also be used to suppress expression of an endogenous transcription factor, thereby reducing or eliminating its activity and modifying one or more traits. Methods for producing such constructs are described in U.S. Patent No. 5,583,021. Preferably, such constructs are made by introducing a premature stop codon into the transcription factor gene. Alternatively, a plant trait can be modified by gene silencing using double-strand RNA (Sharp (1999) Genes and Development 13: 139-141). Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a transcription factor or transcription factor homologue gene. Plants containing a single transgene insertion event at the desired gene can be crossed to generate homozygous plants for the mutation. Such methods are well known to those of skill in the art. (See for example Koncz et al. (1992) Methods in Arabidopsis Research, World Scientific.)

Alternatively, a plant phenotype can be altered by eliminating an endogenous gene, such as a transcription factor or transcription factor homologue, e.g., by homologous recombination (Kempin et al. (1997) Nature 389:802-803).

A plant trait can also be modified by using the Cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome can be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites

is excised. If the lox sites are in the opposite-orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention can also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al. (1997) Nature 390 698-701; Kakimoto et al. (1996) Science 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant can be modified so as to increase transcription levels of a polynucleotide of the invention (See, e.g., PCT Publications WO 96/06166 and WO 98/53057 which describe the modification of the DNA-binding specificity of zinc finger proteins by changing particular amino acids in the DNA-binding motif).

The transgenic plant can also include the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

Transgenic plants (or plant cells, or plant explants, or plant tissues) incorporating the polynucleotides of the invention and/or expressing the polypeptides of the invention can be produced by a variety of well established techniques as described above. Following construction of a vector, most typically an expression cassette, including a polynucleotide, e.g., encoding a transcription factor or transcription factor homologue, of the invention, standard techniques can be used to introduce the polynucleotide into a plant, a plant cell, a plant explant or a plant tissue of interest. Optionally, the plant cell, explant or tissue can be regenerated to produce a transgenic plant.

The plant can be any higher plant, including gymnosperms, monocotyledonous and dicotyledonous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip),

Cruciferae (cabbage, radish, rapeseed, broccoli, etc.), *Curcubitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) Handbook of Plant Cell Culture –Crop Species, Macmillan Publ. Co. Shimamoto et al. (1989) Nature 338:274-276; Fromm et al. (1990) Bio/Technology 8:833-839; and Vasil et al. (1990) Bio/Technology 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods can include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* mediated transformation. Transformation means introducing a nucleotide sequence into a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait can be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in

expression levels or activity of the polypeptide or polynucleotide of the invention can be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

XVII. Integrated Systems – Sequence Identity

Additionally, the present invention may be an integrated system, computer or computer readable medium that comprises an instruction set for determining the identity of one or more sequences in a database. In addition, the instruction set can be used to generate or identify sequences that meet any specified criteria. Furthermore, the instruction set may be used to associate or link certain functional benefits, such improved characteristics, with one or more identified sequence.

For example, the instruction set can include, e.g., a sequence comparison or other alignment program, e.g., an available program such as, for example, the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madison, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PHYTOSEQ sequence database (Incyte Genomics, Palo Alto, CA) can be searched.

Alignment of sequences for comparison can be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443-453, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85:2444-2448, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window can be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al., *supra*.

A variety of methods for determining sequence relationships can be used, including manual alignment and computer assisted sequence alignment and analysis. This later approach is a preferred approach in the present invention, due to the increased throughput afforded by computer assisted methods. As noted above, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (see internet website at ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915). Unless otherwise indicated, "sequence identity" here refers to the % sequence identity generated from a tblastx using the NCBI version of

the algorithm at the default settings using gapped alignments with the filter "off" (see, for example, internet website at ncbi.nlm.nih.gov).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001. An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The program can align, *e.g.,* up to 300 sequences of a maximum length of 5,000 letters.

The integrated system, or computer typically includes a user input interface allowing a user to selectively view one or more sequence records corresponding to the one or more character strings, as well as an instruction set which aligns the one or more character strings with each other or with an additional character string to identify one or more region of sequence similarity. The system may include a link of one or more character strings with a particular phenotype or gene function. Typically, the system includes a user readable output element that displays an alignment produced by the alignment instruction set.

The methods of this invention can be implemented in a localized or distributed computing environment. In a distributed environment, the methods may implemented on a single computer comprising multiple processors or on a multiplicity of computers. The computers can be linked, *e.g.* through a common bus, but more preferably the computer(s) are nodes on a network. The network can be a generalized or a dedicated local or wide-area network and, in certain preferred embodiments, the computers may be components of an intra-net or an internet.

Thus, the invention provides methods for identifying a sequence similar or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

Any sequence herein can be entered into the database, before or after querying the database. This provides for both expansion of the database and, if done before the querying step, for insertion of control sequences into the database. The control sequences can be detected by the query to ensure the general integrity of both the database and the query. As noted, the query can be performed using a web browser based interface. For example, the database can be a centralized public database such as those noted herein, and the querying can be done from a remote terminal or computer across an internet or intranet.

XVIII. Examples

The following examples are intended to illustrate but not limit the present invention. The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4 and Table 6.

Example I: Full Length Gene Identification and Cloning

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of -4 or -5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, *Arabidopsis thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were

synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60°C) and labeled with ^{32}P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M NaPO_4 pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60°C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSSC, 1% SDS at 60°C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the U.C. Marathon cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the U.C. Marathon Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA.

Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process can be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

Example II: Construction of Expression Vectors

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN20 or pMEN65, which are both derived from pMON316 (Sanders et al, (1987) Nucleic Acids Research 15:1543-1558) and contain the CaMV 35S promoter to express transgenes. To clone the sequence into the vector, both pMEN20 and the amplified DNA fragment were digested separately with SalI and NotI restriction enzymes at 37°C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, Valencia CA). The

fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, Beverly MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma, St. Louis, MO). Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen).

Example III: Transformation of *Agrobacterium* with the Expression Vector

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation were made as described by Nagel et al. (1990) FEMS Microbiol Letts. 67: 325-328. *Agrobacterium* strain ABI was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance (A_{600}) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4° C. Cells were then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad, Hercules, CA). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The

presence of the plasmid construct was verified by PCR amplification and sequence analysis.

Example IV: Transformation of *Arabidopsis* Plants with *Agrobacterium tumefaciens* with Expression Vector

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an optical absorbance at 600 nm wavelength over 1 cm (A_{600}) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 μ M benzylamino purine (Sigma), 200 μ l/l Silwet L-77 (Lehle Seeds) until an A_{600} of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 μ E/m²/sec) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

Example V: Identification of *Arabidopsis* Primary Transformants

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H₂O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H₂O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75 $\mu\text{E}/\text{m}^2/\text{sec}$) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T₁ generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants were crossed and progeny seeds (T₂) collected; kanamycin resistant seedlings were selected and analyzed. The expression levels of the recombinant polynucleotides in the transformants varies from about a 5% expression level increase to a least a 100% expression level increase. Similar observations are made with respect to polypeptide level expression.

Example VI: Identification of *Arabidopsis* Plants with Transcription Factor Gene Knockouts

The screening of insertion mutagenized *Arabidopsis* collections for null mutants in a known target gene was essentially as described in Krysan et al (1999) Plant Cell 11:2283-2290. Briefly, gene-specific primers, nested by 5-250 base pairs to each other, were designed from the 5' and 3' regions of a known target gene. Similarly, nested sets of primers were also created specific to each of the T-DNA or

transposon ends (the "right" and "left" borders). All possible combinations of gene specific and T-DNA/transposon primers were used to detect by PCR an insertion event within or close to the target gene. The amplified DNA fragments were then sequenced which allows the precise determination of the T-DNA/transposon insertion point relative to the target gene. Insertion events within the coding or intervening sequence of the genes were deconvoluted from a pool comprising a plurality of insertion events to a single unique mutant plant for functional characterization. The method is described in more detail in Yu and Adam, US Application Serial No. 09/177,733 filed October 23, 1998.

Example VII: Identification of Modified Phenotypes in Overexpression or Gene Knockout Plants

Experiments were performed to identify those transformants or knockouts that exhibited modified biochemical characteristics. Among the biochemicals that were assayed were insoluble sugars, such as arabinose, fucose, galactose, mannose, rhamnose or xylose or the like; prenyl lipids, such as lutein, beta-carotene, xanthophyll-1, xanthophyll-2, chlorophylls A or B, or alpha-, delta- or gamma-tocopherol or the like; fatty acids, such as 16:0 (palmitic acid), 16:1 (palmitoleic acid), 18:0 (stearic acid), 18:1 (oleic acid), 18:2 (linoleic acid), 20:0, 18:3 (linolenic acid), 20:1 (eicosenoic acid), 20:2, 22:1 (erucic acid) or the like; waxes, such as by altering the levels of C29, C31, or C33 alkanes; sterols, such as brassicasterol, campesterol, stigmasterol, sitosterol or stigmastanol or the like, glucosinolates, protein or oil levels.

Fatty acids were measured using two methods depending on whether the tissue was from leaves or seeds. For leaves, lipids were extracted and esterified with hot methanolic H_2SO_4 and partitioned into hexane from methanolic brine. For seed fatty acids, seeds were pulverized and extracted in methanol:heptane:toluene:2,2-dimethoxypropane: H_2SO_4 (39:34:20:5:2) for 90 minutes at 80°C. After cooling to room temperature the upper phase, containing the seed fatty acid esters, was subjected to GC analysis. Fatty acid esters from both seed and leaf tissues were analyzed with a Supelco SP-2330 column.

Glucosinolates were purified from seeds or leaves by first heating the tissue at 95°C for 10 minutes. Preheated ethanol:water (50:50) is and after heating at 95°C for a further 10 minutes, the extraction solvent is applied to a DEAE Sephadex column which had been previously equilibrated with 0.5 M pyridine acetate. Desulfoglucosinolates were eluted with 300 ul water and analyzed by reverse phase HPLC monitoring at 226 nm.

For wax alkanes, samples were extracted using an identical method as fatty acids and extracts were analyzed on a HP 5890 GC coupled with a 5973 MSD. Samples were chromatographically isolated on a J&W DB35 mass spectrometer (J&W Scientific).

To measure prenyl lipids levels, seeds or leaves were pulverized with 1 to 2% pyrogallol as an antioxidant. For seeds, extracted samples were filtered and a portion removed for tocopherol and carotenoid/chlorophyll analysis by HPLC. The remaining material was saponified for sterol determination. For leaves, an aliquot was removed and diluted with methanol and chlorophyll A, chlorophyll B, and total carotenoids measured by spectrophotometry by determining optical absorbance at 665.2 nm, 652.5 nm, and 470 nm. An aliquot was removed for tocopherol and carotenoid/chlorophyll composition by HPLC using a Waters uBondapak C18 column (4.6 mm x 150 mm). The remaining methanolic solution was saponified with 10% KOH at 80°C for one hour. The samples were cooled and diluted with a mixture of methanol and water. A solution of 2% methylene chloride in hexane was mixed in and the samples were centrifuged. The aqueous methanol phase was again re-extracted 2% methylene chloride in hexane and, after centrifugation, the two upper phases were combined and evaporated. 2% methylene chloride in hexane was added to the tubes and the samples were then extracted with one ml of water. The upper phase was removed, dried, and resuspended in 400 ul of 2% methylene chloride in hexane and analyzed by gas chromatography using a 50 m DB-5ms (0.25 mm ID, 0.25 um phase, J&W Scientific).

Insoluble sugar levels were measured by the method essentially described by Reiter et al., (1997) Plant Journal 12:335-345. This method analyzes the neutral sugar composition of cell wall polymers found in *Arabidopsis* leaves. Soluble sugars were

separated from sugar polymers by extracting leaves with hot 70% ethanol. The remaining residue containing the insoluble polysaccharides was then acid hydrolyzed with allose added as an internal standard. Sugar monomers generated by the hydrolysis were then reduced to the corresponding alditols by treatment with NaBH₄, then were acetylated to generate the volatile alditol acetates which were then analyzed by GC-FID. Identity of the peaks was determined by comparing the retention times of known sugars converted to the corresponding alditol acetates with the retention times of peaks from wild-type plant extracts. Alditol acetates were analyzed on a Supelco SP-2330 capillary column (30 m x 250 μ m x 0.2 μ m) using a temperature program beginning at 180° C for 2 minutes followed by an increase to 220° C in 4 minutes. After holding at 220° C for 10 minutes, the oven temperature is increased to 240° C in 2 minutes and held at this temperature for 10 minutes and brought back to room temperature.

To identify plants with alterations in total seed oil or protein content, 150mg of seeds from T2 progeny plants were subjected to analysis by Near Infrared Reflectance Spectroscopy (NIRS) using a Foss NirSystems Model 6500 with a spinning cup transport system. NIRS is a non-destructive analytical method used to determine seed oil and protein composition. Infrared is the region of the electromagnetic spectrum located after the visible region in the direction of longer wavelengths. 'Near infrared' owns its name for being the infrared region near to the visible region of the electromagnetic spectrum. For practical purposes, near infrared comprises wavelengths between 800 and 2500 nm. NIRS is applied to organic compounds rich in O-H bonds (such as moisture, carbohydrates, and fats), C-H bonds (such as organic compounds and petroleum derivatives), and N-H bonds (such as proteins and amino acids). The NIRS analytical instruments operate by statistically correlating NIRS signals at several wavelengths with the characteristic or property intended to be measured. All biological substances contain thousands of C-H, O-H, and N-H bonds. Therefore, the exposure to near infrared radiation of a biological sample, such as a seed, results in a complex spectrum which contains qualitative and quantitative information about the physical and chemical composition of that sample.

The numerical value of a specific analyte in the sample, such as protein content or oil content, is mediated by a calibration approach known as chemometrics.

Chemometrics applies statistical methods such as multiple linear regression (MLR), partial least squares (PLS), and principle component analysis (PCA) to the spectral data and correlates them with a physical property or other factor, that property or factor is directly determined rather than the analyte concentration itself. The method first provides "wet chemistry" data of the samples required to develop the calibration.

Calibration for Arabidopsis seed oil composition was performed using accelerated solvent extraction using 1 g seed sample size and was validated against certified canola seed. A similar wet chemistry approach was performed for seed protein composition calibration.

Data obtained from NIRS analysis was analyzed statistically using a nearest-neighbor (N-N) analysis. The N-N analysis allows removal of within-block spatial variability in a fairly flexible fashion which does not require prior knowledge of the pattern of variability in the chamber. Ideally, all hybrids are grown under identical experimental conditions within a block (rep). In reality, even in many block designs, significant within-block variability exists. Nearest-neighbor procedures are based on assumption that environmental effect of a plot is closely related to that of its neighbors. Nearest-neighbor methods use information from adjacent plots to adjust for within-block heterogeneity and so provide more precise estimates of treatment means and differences. If there is within-plot heterogeneity on a spatial scale that is larger than a single plot and smaller than the entire block, then yields from adjacent plots will be positively correlated. Information from neighboring plots can be used to reduce or remove the unwanted effect of the spatial heterogeneity, and hence improve the estimate of the treatment effect. Data from neighboring plots can also be used to reduce the influence of competition between adjacent plots. The Papadakis N-N analysis can be used with designs to remove within-block variability that would not be removed with the standard split plot analysis (Papadakis, 1973, Inst. d'Amelior. Plantes Thessaloniki (Greece) Bull. Scientif., No. 23; Papadakis, 1984, Proc. Acad. Athens, 59, 326-342).

Experiments were performed to identify those transformants or knockouts that exhibited an improved pathogen tolerance. For such studies, the transformants were exposed to biotrophic fungal pathogens, such as *Erysiphe orontii*, and necrotrophic

fungus pathogens, such as *Fusarium oxysporum*. *Fusarium oxysporum* isolates cause vascular wilts and damping off of various annual vegetables, perennials and weeds (Mauch-Mani and Slusarenko (1994) Molecular Plant-Microbe Interactions 7: 378-383). For *Fusarium oxysporum* experiments, plants grown on Petri dishes were sprayed with a fresh spore suspension of *F. oxysporum*. The spore suspension was prepared as follows: A plug of fungal hyphae from a plate culture was placed on a fresh potato dextrose agar plate and allowed to spread for one week. 5 ml sterile water was then added to the plate, swirled, and pipetted into 50 ml Armstrong *Fusarium* medium. Spores were grown overnight in *Fusarium* medium and then sprayed onto plants using a Preval paint sprayer. Plant tissue was harvested and frozen in liquid nitrogen 48 hours post infection.

Erysiphe orontii is a causal agent of powdery mildew. For *Erysiphe orontii* experiments, plants were grown approximately 4 weeks in a greenhouse under 12 hour light (20°C, ~30% relative humidity (rh)). Individual leaves were infected with *E. orontii* spores from infected plants using a camel's hair brush, and the plants were transferred to a Percival growth chamber (20°C, 80% rh.). Plant tissue was harvested and frozen in liquid nitrogen 7 days post infection.

Botrytis cinerea is a necrotrophic pathogen. *Botrytis cinerea* was grown on potato dextrose agar in the light. A spore culture was made by spreading 10 ml of sterile water on the fungus plate, swirling and transferring spores to 10 ml of sterile water. The spore inoculum (approx. 105 spores/ml) was used to spray 10 day-old seedlings grown under sterile conditions on MS (minus sucrose) media. Symptoms were evaluated every day up to approximately 1 week.

Infection with bacterial pathogens *Pseudomonas syringae* pv *maculicola* (Psm) strain 4326 and pv *maculicola* strain 4326 was performed by hand inoculation at two doses. Two inoculation doses allows the differentiation between plants with enhanced susceptibility and plants with enhanced resistance to the pathogen. Plants were grown for 3 weeks in the greenhouse, then transferred to the growth chamber for the remainder of their growth. Psm ES4326 was hand inoculated with 1 ml syringe on 3 fully-expanded leaves per plant (4 1/2 wk old), using at least 9 plants per overexpressing line at two inoculation doses, OD=0.005 and OD=0.0005. Disease

scoring occurred at day 3 post-inoculation with pictures of the plants and leaves taken in parallel.

In some instances, expression patterns of the pathogen-induced genes (such as defense genes) was monitored by microarray experiments. cDNAs were generated by PCR and resuspended at a final concentration of ~ 100 ng/ul in 3X SSC or 150mM Na-phosphate (Eisen and Brown (1999) *Methods Enzymol.* 303:179-205). The cDNAs were spotted on microscope glass slides coated with polylysine. The prepared cDNAs were aliquoted into 384 well plates and spotted on the slides using an x-y-z gantry (OmniGrid) purchased from GeneMachines (Menlo Park, CA) outfitted with quill type pins purchased from Telechem International (Sunnyvale, CA). After spotting, the arrays were cured for a minimum of one week at room temperature, rehydrated and blocked following the protocol recommended by Eisen and Brown (1999; *supra*).

Sample total RNA (10 ug) samples were labeled using fluorescent Cy3 and Cy5 dyes. Labeled samples were resuspended in 4X SSC/0.03% SDS/4 ug salmon sperm DNA/2 ug tRNA/ 50mM Na-pyrophosphate, heated for 95°C for 2.5 minutes, spun down and placed on the array. The array was then covered with a glass coverslip and placed in a sealed chamber. The chamber was then kept in a water bath at 62°C overnight. The arrays were washed as described in Eisen and Brown (1999) and scanned on a General Scanning 3000 laser scanner. The resulting files are subsequently quantified using Imagene, a software purchased from BioDiscovery (Los Angeles, CA).

Experiments were performed to identify those transformants or knockouts that exhibited an improved environmental stress tolerance. For such studies, the transformants were exposed to a variety of environmental stresses. Plants were exposed to chilling stress (6 hour exposure to 4-8° C), heat stress (6 hour exposure to 32-37° C), high salt stress (6 hour exposure to 200 mM NaCl), drought stress (168 hours after removing water from trays), osmotic stress (6 hour exposure to 3 M mannitol), or nutrient limitation (nitrogen, phosphate, and potassium) (Nitrogen: all components of MS medium remained constant except N was reduced to 20 mg/l of NH₄NO₃, or Phosphate: All components of MS medium except KH₂PO₄, which was

replaced by K_2SO_4 , Potassium: All components of MS medium except removal of KNO_3 and KH_2PO_4 , which were replaced by NaH_4PO_4).

Experiments were performed to identify those transformants or knockouts that exhibited a modified structure and development characteristics. For such studies, the transformants were observed by eye to identify novel structural or developmental characteristics associated with the ectopic expression of the polynucleotides or polypeptides of the invention.

Experiments were performed to identify those transformants or knockouts that exhibited modified sugar-sensing. For such studies, seeds from transformants were germinated on media containing 5% glucose or 9.4% sucrose which normally partially restrict hypocotyl elongation. Plants with altered sugar sensing may have either longer or shorter hypocotyls than normal plants when grown on this media. Additionally, other plant traits may be varied such as root mass.

Flowering time was measured by the number of rosette leaves present when a visible inflorescence of approximately 3 cm is apparent. Rosette and total leaf number on the progeny stem are tightly correlated with the timing of flowering (Koornneef et al (1991) *Mol. Gen. Genet* 229:57-66). The vernalization response was measured. For vernalization treatments, seeds were sown to MS agar plates, sealed with micropore tape, and placed in a 4°C cold room with low light levels for 6-8 weeks. The plates were then transferred to the growth rooms alongside plates containing freshly sown non-vernalized controls. Rosette leaves were counted when a visible inflorescence of approximately 3 cm was apparent.

Modified phenotypes observed for particular overexpressor or knockout plants are provided in Table 4. For a particular overexpressor that shows a less beneficial characteristic, it may be more useful to select a plant with a decreased expression of the particular transcription factor. For a particular knockout that shows a less beneficial characteristic, it may be more useful to select a plant with an increased expression of the particular transcription factor.

The sequences of the Sequence Listing or those in Tables 4 , 5 or those disclosed here can be used to prepare transgenic plants and plants with altered traits. The specific transgenic plants listed below are produced from the sequences of the Sequence Listing, as noted. Table 4 provides exemplary polynucleotide and polypeptide sequences of the invention. Table 4 includes, from left to right for each sequence: the first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Seed of plants overexpressing sequences G265 (SEQ ID NOs:871 and 872), G715 (SEQ ID NOs:925 and 926), G1471 (SEQ ID NOs:311 and 312), G1793 (SEQ ID NOs:365 and 366), G1838 (SEQ ID NOs:381 and 382), G1902 (SEQ ID NOs:405 and 406), G286 (SEQ ID NOs:877 and 878), G2138 (SEQ ID NOs:865 and 866) and G2830 (SEQ ID NOs:875 and 876) was subjected to NIR analysis and a significant increase in seed oil content compared with seed from control plants was identified.

G192: G192 (SEQ ID NO: 859) was expressed in all plant tissues and under all conditions examined. Its expression was slightly induced upon infection by *Fusarium*. G192 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G192 overexpressors were late flowering under 12 hour light and had more leaves than control plants. This phenotype was manifested in the three T2 lines analyzed. Results of one experiment suggest that G192 overexpressor was more susceptible to infection with a moderate dose of the fungal pathogen *Erysiphe orontii*. The decrease in seed oil observed for one line was replicated in an independent experiment. G192 overexpression delayed flowering. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering, or for systems of inducible flowering time control. In particular, in species where the vegetative parts of the plants constitute the crop and

the reproductive tissues are discarded, it will be advantageous to delay or prevent flowering. Extending vegetative development can bring about large increases in yields. G192 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G192 can be used to manipulate seed oil content, which can be of nutritional value.

Closely Related Genes from Other Species

G192 had some similarity within the conserved WRKY domain to non-Arabidopsis plant proteins.

G1946: G1946 (SEQ ID NO: 801) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G1946 resulted in accelerated flowering, with 35S::G1946 transformants producing flower buds up to a week earlier than wild-type controls (24-hour light conditions). These effects were seen in 12/20 primary transformants and in two independent plantings of each of the three T2 lines. Unlike many early flowering Arabidopsis transgenic lines, which are dwarfed, 35S::G1946 transformants often reached full-size at maturity, and produced large quantities of seeds, although the plants were slightly pale in coloration and had slightly flat leaves compared to wild-type. In addition, 35S::G1946 plants showed an altered response to phosphate deprivation. Seedlings of G1946 overexpressor plants showed more secondary root growth on phosphate-free media, when compared to wild-type control. In a repeat experiment, all three lines showed the phenotype. Overexpression of G1946 in Arabidopsis also resulted in an increase in seed glucosinolate M39501 in T2 lines 1 and 3. An increase in seed oil and a decrease in seed protein was also observed in these two lines. G1946 was ubiquitously expressed, and does not appear to be significantly induced or repressed by any of the biotic and abiotic stress conditions tested at this time, with the exception of cold, which repressed G1946 expression. G1946 can be used to modify flowering time, as well as to improve the plant's performance in conditions of limited phosphate, and to alter seed oil, protein, and glucosinolate composition.

Closely Related Genes from Other Species

A comparison of the amino acid sequence of G1946 with sequences available from GenBank showed strong similarity with plant HSFs of several species (*Lycopersicon peruvianum*, *Medicago truncatula*, *Lycopersicon esculentum*, *Glycine max*, *Solanum tuberosum*, *Oryza sativa* and *Hordeum vulgare* subsp. *vulgare*).

G375: The sequence of G375 (SEQ ID NO:239) was experimentally determined and G375 was analyzed using transgenic plants in which G375 was expressed under the control of the 35S promoter. Overexpression of G375 produced marked effects on leaf development. At early stages of growth, 35S::G375 seedlings developed narrow, upward pointing leaves with long petioles (possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements). Additionally, some seedlings were noted to have elongated hypocotyls, and some were rather small compared to wild-type controls. Comparable phenotypes were obtained by overexpression of an AP2 family gene, G2113 (SEQ ID NO: 85). Following the switch to flowering, 35S::G375 plants showed reduced fertility, which possibly arose from a failure of stamens to fully elongate. One of the three T2 lines, (#41) was later flowering than wild-type controls, and also developed large numbers of small secondary rosette leaves in the axils of the primary rosette. Although these effects were not noted in the other two lines, the phenotypes obtained in line 41 were somewhat similar to those produced by overexpression of another Z-dof gene, G736 (SEQ ID NO: 211). G375 was expressed in all tissues, although at different levels. It was expressed at low levels in the root and germinating seed, and expressed at high levels in the embryo. The effects of G375 on leaf architecture are of potential interest to the ornamental horticulture industry.

Closely Related Genes from Other Species

G375 showed some homology to non-Arabidopsis plant proteins within the conserved Dof domain.

G1255: The sequence of G1255 (SEQ ID NO: 273) was experimentally determined and G1255 was analyzed using transgenic plants in which G1255 was expressed under the control of the 35S promoter. Plants overexpressing G1255 had

alterations in leaf architecture, a reduction in apical dominance, an increase in seed size, and showed more disease symptoms following inoculation with a low dose of the fungal pathogen *Botrytis cinerea*. G1255 was constitutively expressed and not significantly induced by any conditions tested. On the basis of the phenotypes produced by overexpression of G1255, G1255 can be used to manipulate the plant's defense response to produce pathogen resistance, alter plant architecture, or alter seed size.

Closely Related Genes from Other Species

G1255 showed strong homology to a putative rice zing finger protein represented by sequence AC087181_3. Sequence identity between these two protein extended beyond the conserved domain, and therefore, these genes can be orthologs.

G865: The complete cDNA sequence of G865 (SEQ ID NO: 557) was determined. G865 was ubiquitously expressed in *Arabidopsis* tissues. G865 was analyzed using transgenic plants in which G865 was expressed under the control of the 35S promoter. Plants overexpressing G865 were early flowering, with numerous secondary inflorescence meristems giving them a bushy appearance. G865 overexpressors were more susceptible to infection with a moderate dose of the fungal pathogens *Erysiphe orontii* and *Botrytis cinerea*. In addition, seeds from G865 overexpressing plants showed a trend of increased protein and reduced oil content, although the observed changes were not beyond the criteria used for judging significance except in one line. G865 can be used to control flowering time. G865 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G865 can be used to alter seed oil and protein content of a plant.

Closely Related Genes from Other Species

G865 and other non-*Arabidopsis* AP2/EREBP proteins were similar within the conserved AP2 domain.

G2509: G2509 (SEQ ID NO: 23) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G2509 caused multiple alterations in plant growth and development, most notably, altered branching patterns, and a reduction in apical dominance, giving the plants a

shorter, more bushy stature than wild type. Twenty 35S::G2509 primary transformants were examined; at early stages of rosette development, these plants displayed a wild-type phenotype. However, at the switch to flowering, almost all T1 lines showed a marked loss of apical dominance and large numbers of secondary shoots developed from axils of primary rosette leaves. In the most extreme cases, the shoots had very short internodes, giving the inflorescence a very bushy appearance. Such shoots were often very thin and flowers were relatively small and poorly fertile. At later stages, many plants appeared very small and had a low seed yield compared to wild type. In addition to the effects on branching, a substantial number of 35S::G2509 primary transformants also flowered early and had buds visible several days prior to wild type. Similar effects on inflorescence development were noted in each of three T2 populations examined. The branching and plant architecture phenotypes observed in 35S::G2509 lines resemble phenotypes observed for three other AP2/EREBP genes: G865 (SEQ ID NO: 557), G1411 (SEQ ID NO: 3), and G1794 (SEQ ID NO: 13). G2509, G865, and G1411 form a small clade within the large AP2/EREBP family, and G1794, although not belonging to the clade, is one of the AP2/EREBP genes closest to it in the phylogenetic tree. It is thus likely that all these genes share a related function, such as affecting hormone balance. Overexpression of G2509 in Arabidopsis resulted in an increase in alpha-tocopherol in seeds in T2 lines 5 and 11. G2509 was ubiquitously expressed in Arabidopsis plant tissue. G2509 expression levels were altered by a variety of environmental or physiological conditions. G2509 can be used to manipulate plant architecture and development. G2509 can be used to alter tocopherol composition. Tocopherols have anti-oxidant and vitamin E activity. G2509 can be useful in altering flowering time. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G2509 showed some sequence similarity with known genes from other plant species within the conserved AP2/EREBP domain.

G2347: G2347 (SEQ ID NO: 1119) was analyzed using transgenic plants in which G2347 was expressed under the control of the 35S promoter. Overexpression of G2347 markedly reduced the time to flowering in Arabidopsis. This phenotype

was apparent in the majority of primary transformants and in all plants from two out of the three T2 lines examined. Under continuous light conditions, 35S::G2347 plants formed flower buds up a week earlier than wild type. Many of the plants were rather small and spindly compared to controls. To demonstrate that overexpression of G2347 could induce flowering under less inductive photoperiods, two T2 lines were re-grown in 12 hour conditions; again, all plants from both lines bolted early, with some initiating flower buds up to two weeks sooner than wild-type. As determined by RT-PCR, G2347 was highly expressed in rosette leaves and flowers, and to much lower levels in embryos and siliques. No expression of G2347 was detected in the other tissues tested. G2347 expression was repressed by cold, and by auxin treatments and by infection by Erysiphe. G2347 is also highly similar to the Arabidopsis protein G2010 (SEQ ID NO: 1121). The level of homology between these two proteins suggested they could have similar, overlapping, or redundant functions in Arabidopsis. In support of this hypothesis, overexpression of both G2010 and G2347 resulted in early flowering phenotypes in transgenic plants.

Closely Related Genes from Other Species

The closest relative to G2347 is the Antirrhinum protein, SBP2 (CAA63061). The similarity between these two proteins is extensive enough to suggest they might have similar functions in a plant.

G988: G988 (SEQ ID NO: 43) was analyzed using transgenic plants in which G988 was expressed under the control of the 35S promoter. Plants overexpressing G988 had multiple morphological phenotypes. The transgenic plants were generally smaller than wild-type plants, had altered leaf, inflorescence and flower development, altered plant architecture, and altered vasculature. In one transgenic line overexpressing G988 (line 23), an increase in the seed glucosinolate M39489 was observed. The phenotype of plants overexpressing G988 was wild-type in all other assays performed. In wild-type plants, G988 was expressed primarily in flower and silique tissue, but was also present at detectable levels in all other tissues tested. Expression of G988 was induced in response to heat treatment, and repressed in response to infection with Erysiphe. Based on the observed morphological phenotypes of the transgenic plants, G988 can be used to create plants with larger flowers. This can have value in the ornamental horticulture industry. The reduction

in the formation of lateral branches suggests that G988 can have utility on the forestry industry. The Arabidopsis plants overexpressing G988 also had reduced fertility. This can be a desirable trait in some instances, as it can be exploited to prevent or minimize the escape of GMO (genetically modified organism) pollen into the environment.

Closely Related Genes from Other Species

The amino acid sequence for the *Capsella rubella* hypothetical protein represented by GenBank accession number CRU303349 was significantly identical to G988 outside of the SCR conserved domains. The *Capsella rubella* hypothetical protein is 90% identical to G988 over a stretch of roughly 450 amino acids. Therefore, it is likely that the *Capsella rubella* gene is an ortholog of G988.

G2346: G2346 (SEQ ID NO: 459) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. 35S::G2346 seedlings from all three T2 populations had slightly larger cotyledons and appeared somewhat more advanced than controls. This indicated that the seedlings developed more rapidly than the control plants. At later stages, however, G2346 overexpressing plants showed no consistent differences from control plants. The phenotype of these transgenic plants was wild-type in all other assays performed. According to RT-PCR analysis, G2346 is expressed ubiquitously.

Closely Related Genes from Other Species

G2346 shows some sequence similarity with known genes from other plant species within the conserved SBP domain.

G1354: The complete sequence of G1354 (SEQ ID NO: 285) was determined. G1354 was analyzed using transgenic plants in which G1354 was expressed under the control of the 35S promoter. Overexpression of G1354 produced highly deleterious effects on growth and development. Only three 35S::G1354 T1 plants were obtained; all were extremely tiny and slow developing. After three weeks of growth, each of the plants comprised a completely disorganized mass of leaves and root that had no clear axis of growth. Since these individuals would not have survived transplantation to soil, they were harvested for RT-PCR analysis; all three plants showed moderate

levels of G1354 overexpression compared to whole wild-type seedlings of an equivalent size. Only a very small number of transformants were obtained from two selection attempts on separate batches of T0 seed. Usually between 15 and 120 transformants are obtained from each aliquot of 300 mg T0 seed from wild-type plants. The low transformation frequency obtained in this experiment suggests that high levels of G1354 overexpression might have completely lethal effects and prevent transformed seeds from germinating. As determined by RT-PCR, G1354 was uniformly expressed in all tissues and under all conditions tested in RT-PCR. However, the gene was repressed in leaf tissue in response to Erysiphe infection.

Closely Related Genes from Other Species

G1354 is closely related to a NAM protein encoded by polynucleotide from rice (AC005310). Similarity between G1354 and this rice protein extends beyond the signature motif of the family to a level that would suggest the genes are orthologs.

G1063: G1063 (SEQ ID NO: 119) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1499 (SEQ ID NO: 7), G2143 (SEQ ID NO: 129), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. A spectrum of developmental alterations was observed amongst 35S::G1063 primary transformants and the majority were markedly small, dark green, and had narrow curled leaves. The most severely affected individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures. In other cases, flowers showed internode elongation between floral whorls, with a central carpel protruding on a pedicel-like organ. Additionally, lateral branches sometimes failed to develop and tiny patches of carpelloid tissue formed at axillary nodes of the inflorescence. In lines with an intermediate phenotype, flowers contained defined whorls of organs, but sepals were converted to carpelloid structures or displayed patches of carpelloid tissue. In contrast, lines with a weak phenotype developed relatively normal flowers and produced a reasonable quantity of seed. Such plants were still distinctly smaller than wild-type controls. Since the strongest 35S::G1063 lines were sterile, three lines

with a relatively weak phenotype, that had produced sufficient seed for biochemical and physiological analysis, were selected for further study. Two of the T2 populations (T2-28,37) were clearly small, darker green and possessed narrow leaves compared to wild type. Plants from one of these populations (T2-28) also produced occasional branches with abnormal flowers like those seen in the T1. The final T2 population (T2-30) displayed a very mild phenotype. Overexpression of G1063 in *Arabidopsis* resulted in a decrease in seed oil content in T2 lines 28 and 37. No altered phenotypes were detected in any of the physiological assays, except that the plants were noted to be somewhat small and produce anthocyanin when grown in Petri plates. G1063 was expressed at low to moderate levels in roots, flowers, rosette leaves, embryos, and germinating seeds, but was not detected in shoots or siliques. It was induced by auxin. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*. G1063 has utility in manipulating seed oil and protein content.

Closely Related Genes from Other Species

G1063 protein shared extensive homology in the basic helix loop helix region with a protein sequence encoded by Glycine max cDNA clone (AW832545) as well as a tomato root, plants pre-anthesis *Lycopersicon esculentum* cDNA (BE451174).

G2143: G2143 (SEQ ID NO: 129) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. Twelve out of twenty 35S::G2143 T1 lines showed a very severe phenotype; these plants were markedly small and had narrow, curled, dark-green leaves. Such individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures, or a fused mass of carpelloid tissue. Furthermore, lateral branches usually failed to develop, and tiny patches of stigmatic tissue often formed at axillary nodes of the inflorescence. Strongly affected plants displayed the highest levels of transgene expression

(determined by RT-PCR). The remaining T1 lines showed lower levels of G2143 overexpression; these plants were still distinctly smaller than wild type, but had relatively normal inflorescences and produced seed. Since the strongest 35S::G2143 lines were sterile, three lines with a relatively weak phenotype, that had produced sufficient seed for biochemical analysis, were selected for further study. T2-11 plants displayed a very mild phenotype and had somewhat small, narrow, dark green leaves. The other two T2 populations, however, appeared wild-type, suggesting that transgene activity might have been reduced between the generations. Reduced seedling vigor was noted in the physiological assays. G2143 expression was detected at low levels in flowers and siliques, and at higher levels in germinating seed. G2143 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2143 protein shared extensive homology in the basic helix loop helix region with a protein encoded by Glycine max cDNA clones (AW832545, BG726819 and BG154493) and a *Lycopersicon esculentum* cDNA clone (BE451174). There was lower homology outside of the region.

G2557: G2557 (SEQ ID NO: 133) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2143 (SEQ ID NO: 129). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. The flowers of 35S::G2557 primary transformants displayed patches of stigmatic papillae on the sepals, and often had rather narrow petals and poorly developed stamens. Additionally, carpels were also occasionally held outside of the flower at the end of an elongated pedicel like structure. As a result of such defects, 35S::G2557 plants often showed very poor fertility and formed small wrinkled siliques. In addition to such floral abnormalities, the majority of primary transformants were also small and darker green in coloration than wild type. Approximately one third of the T1 plants were extremely tiny and completely sterile. Three T1 lines (#7,9,12), that had produced some seeds, and

showed a relatively weak phenotype, were chosen for further study. All three of the T2 populations from these lines contained plants that were distinctly small, had abnormal flowers, and were poorly fertile compared to controls. Stigmatic tissue was not noted on the sepals of plants from these three T2 lines. Another line (#4) that had shown a moderately strong phenotype in the T1 was sown for only morphological analysis in the T2 generation. T2-4 plants were small, dark green, and produced abnormal flowers with ectopic stigmatic tissue on the sepals, as had been seen in the parental plant. G2557 expression was detected at low to moderate levels in all tissues tested except shoots. It was induced by cold, heat, and salt, and repressed by pathogen infection. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2557 protein shows extensive sequence similarity in the region of basic helix loop helix with a protein encoded by Glycine max cDNA clone (BE347811).

G2430: The complete sequence of G2430 (SEQ ID NO: 697) was determined. G2430 is a member of the response regulator class of GARP proteins (ARR genes), although one of the two conserved aspartate residues characteristic of response regulators is not present. The second aspartate, the putative phosphorylated site, is retained so G2430 can have response regulator function. G2430 is specifically expressed in embryo and silique tissue. In morphological analyses, plants overexpressing G2430 showed more rapid growth than control plants at early stages, and in two of three lines examined produced large, flat leaves. Early flowering was observed for some lines, but this effect was inconsistent between plantings. G2430 can regulate plant growth. Overexpression of G2430 in *Arabidopsis* also resulted in seedlings that are slightly more tolerant to heat in a germination assay. Seedlings from G2430 overexpressing transgenic plants were slightly greener than the control seedlings under high temperature conditions. In a repeat experiment on individual lines, G2430 line 15 showed the strongest heat tolerant phenotype. G2430 can be useful to promote faster development and reproduction in plants.

Closely Related Genes from Other Species

G2430 had some similarity within of the conserved GARP and response-regulator domains to non-Arabidopsis proteins.

G1478: The sequence of G1478 (SEQ ID NO: 831) was determined and G1478 was analyzed using transgenic plants in which G1478 was expressed under the control of the 35S promoter. Plants overexpressing G1478 had a general delay in progression through the life cycle, in particular a delay in flowering time. G1478 is expressed at higher levels in flowers, rosettes and embryos but otherwise expression is constitutive. Based on the phenotypes produced through G1478 overexpression, G1478 can be used to manipulate the rate at which plants grow, and flowering time.

Closely Related Genes from Other Species

G1478 shows some homology to non-Arabidopsis proteins within the conserved domain.

G681: G681 (SEQ ID NO: 579) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. Approximately half of the 35S::G681 primary transformants were markedly small and formed narrow leaves compared to controls. These plants often produced thin inflorescence stems, had rather poorly formed flowers with low pollen production, and set few seeds. Three T1 lines with relatively weak phenotypes, which had produced reasonable quantities of seed, were selected for further study. Plants from one of the T2 populations were noted to be slightly small, but otherwise the T2 lines displayed no consistent differences in morphology from controls. In leaves of two of the T2 lines, overexpression of G681 resulted in an increase in the percentage of the glucosinolate M39480. According to RT-PCR analysis, G681 expression was detected at very low levels in flower and rosette leaf tissues. G681 was induced by drought stress. G681 can be used to alter glucosinolate composition in plants. Increases or decreases in specific glucosinolates or total glucosinolate content are desirable depending upon the particular application. For example: (1) Glucosinolates are undesirable components of the oilseeds used in animal feed, since they produce toxic effects. Low-glucosinolate varieties of canola have been developed to combat this problem. (2) Some glucosinolates have anti-cancer activity; thus, increasing the levels or

composition of these compounds might be of interest from a nutraceutical standpoint. (3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters can be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Closely Related Genes from Other Species

G681 shows some sequence similarity with known genes from other plant species within the conserved Myb domain.

G878: G878 (SEQ ID NO: 611) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Analysis of primary transformants revealed that overexpression of G878 delays the onset of flowering in Arabidopsis. 11/20 of the 35S::G878 T1 plants flowered approximately one week later than wild type under continuous light conditions. These plants were also darker green, had shorter stems, and senesced later than controls. G878 was ubiquitously expressed. G878 can be used to modify flowering time and senescence, and a wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G878 was highly related to other WRKY proteins from a variety of plant species, such as the *Nicotiana tabacum* DNA-binding protein 2 (WRKY2) (AF096299), and a *Cucumis sativus* SPF1-like DNA-binding protein (L44134).

G374: G374 (SEQ ID NO: 47) was expressed at low levels throughout the plant and was induced by salicylic acid. G374 was investigated using lines carrying a T-DNA insertion in this gene. The T-DNA insertion was approximately three quarters of the way into the protein coding sequence and should result in a null mutation. Homozygosity for a T-DNA insertion within G374 caused lethality at early stages of embryo development. In an initial screen for G374 knockouts, heterozygous plants were identified. Seed from those individuals was sown to soil and eleven plants were PCR-screened to identify homozygotes. No homozygotes were obtained;

6 of the progeny were heterozygous whilst the other 5 were wild type. This raised the prospect that homozygosity for the G374 insertion was lethal. To examine this possibility further, heterozygous KO.G374 plants were re-grown. These individuals looked wild type, but their siliques were examined for seed abnormalities. When green siliques were dissected, around 25% of developing seeds were white or aborted. Embryos from these siliques were cleared using Hoyers solution, and examined under the microscope. It was apparent that embryos from the white seeds had arrested at early (globular or heart) stages of development, whilst embryos from the normal seeds were fully developed. Such arrested or aborted seeds most likely represented homozygotes for the G374 insertion. To support this conclusion, seed was collected from heterozygous plants and sown to kanamycin plates (the T-DNA insertion carried the NPT marker gene). Of the seedlings that germinated, 160 were kanamycin resistant and 107 were kanamycin sensitive. These data more closely fitted a 2:1 (chi-sq., 1df, = 5.5, $0.05 > P > 0.01$) than a 3:1 (chi-sq., 1df, = 32, $P < 0.001$) ratio. Such a segregation ratio suggested that a homozygous class of kanamycin resistant seedlings was absent from the progeny of KO.G374 plant. G374 can be a herbicide target.

Closely Related Genes from Other Species

Similar sequences to G374 are present in tomato and *Medicago truncatula*, and these sequences can be orthologs.

Example VIII: Identification of Homologous Sequences

Homologous sequences from *Arabidopsis* and plant species other than *Arabidopsis* were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; and Altschul et al. (1997) *Nucl. Acid Res.* 25: 3389-3402). The tblastx sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919).

Identified non-*Arabidopsis* sequences homologous to the *Arabidopsis* sequences are provided in Table 5. The percent sequence identity among these sequences can be as low as 47%, or even lower sequence identity. The entire NCBI GenBank database was filtered for sequences from all plants except *Arabidopsis thaliana* by selecting all entries in the NCBI GenBank database associated with NCBI

taxonomic ID 33090 (Viridiplantae; all plants) and excluding entries associated with taxonomic ID 3701 (*Arabidopsis thaliana*). These sequences are compared to sequences representing genes of SEQ IDs NOs:2 - 2N, where $N = 2-561$, using the Washington University TBLASTX algorithm (version 2.0a19MP) at the default settings using gapped alignments with the filter "off". For each gene of SEQ IDs NOs:2 - 2N, where $N = 2-561$, individual comparisons were ordered by probability score (P-value), where the score reflects the probability that a particular alignment occurred by chance. For example, a score of $3.6e-40$ is 3.6×10^{-40} . In addition to P-values, comparisons were also scored by percentage identity. Percentage identity reflects the degree to which two segments of DNA or protein are identical over a particular length. Examples of sequences so identified are presented in Table 5. Homologous or orthologous sequences are readily identified and available in GenBank by Accession number (Table 5; Test sequence ID). The identified homologous polynucleotide and polypeptide sequences and homologues of the *Arabidopsis* polynucleotides and polypeptides may be orthologs of the *Arabidopsis* polynucleotides and polypeptides (TBD: to be determined).

Example IX Introduction of polynucleotides into dicotyledonous plants

SEQ ID NOs:1-(2N - 1), wherein $N = 2-561$, paralogous, orthologous, and homologous sequences recombined into pMEN20 or pMEN65 expression vectors are transformed into a plant for the purpose of modifying plant traits. The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants using most dicot plants (see Weissbach and Weissbach, (1989) *supra*; Gelvin et al., (1990) *supra*; Herrera-Estrella et al. (1983) *supra*; Bevan (1984) *supra*; and Klee (1985) *supra*). Methods for analysis of traits are routine in the art and examples are disclosed above.

Example X Transformation of Cereal Plants with an Expression Vector

Cereal plants such as corn, wheat, rice, sorghum or barley, may also be transformed with the present polynucleotide sequences in pMEN20 or pMEN65 expression vectors for the purpose of modifying plant traits. For example, pMEN020 may be modified to replace the NptII coding region with the BAR gene of *Streptomyces hygroscopicus* that confers resistance to phosphinothricin. The KpnI

and BglII sites of the Bar gene are removed by site-directed mutagenesis with silent codon changes.

The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants of most cereal crops (Vasil, I., Plant Molec. Biol. 25: 925-937 (1994)) such as corn, wheat, rice, sorghum (Cassas, A. et al., Proc. Natl. Acad Sci USA 90: 11212-11216 (1993) and barley (Wan, Y. and Lemeaux, P. Plant Physiol. 104:37-48 (1994). DNA transfer methods such as the microprojectile can be used for corn (Fromm, et al. Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al. Plant Cell 2: 603-618 (1990); Ishida, Y., Nature Biotechnology 14:745-750 (1990)), wheat (Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), rice (Christou Bio/Technology 9:957-962 (1991); Hiei et al. Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617; Hiei et al., Plant Mol Biol. 35:205-18 (1997)). For most cereal plants, embryogenic cells derived from immature scutellum tissues are the preferred cellular targets for transformation (Hiei et al., Plant Mol Biol. 35:205-18 (1997); Vasil, Plant Molec. Biol. 25: 925-937 (1994)).

Vectors according to the present invention may be transformed into corn embryogenic cells derived from immature scutellar tissue by using microprojectile bombardment, with the A188XB73 genotype as the preferred genotype (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). After microprojectile bombardment the tissues are selected on phosphinothricin to identify the transgenic embryogenic cells (Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). Transgenic plants are regenerated by standard corn regeneration techniques (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)).

The plasmids prepared as described above can also be used to produce transgenic wheat and rice plants (Christou, Bio/Technology 9:957-962 (1991); Hiei et al., Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617 (1996); Hiei et al., Plant Mol Biol. 35:205-18 (1997)) that coordinately express genes of

interest by following standard transformation protocols known to those skilled in the art for rice and wheat Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), where the bar gene is used as the selectable marker.

All references, publications, patent documents, web pages, and other documents cited or mentioned herein are hereby incorporated by reference in their entirety for all purposes. Although the invention has been described with reference to specific embodiments and examples, it should be understood that one of ordinary skill can make various modifications without departing from the spirit of the invention. The scope of the invention is not limited to the specific embodiments and examples provided.

We claim:

1. A transgenic plant comprising a recombinant polynucleotide having a nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from those of SEQ ID NOs: 860, 802, 240, 274, 558, 24, 1120, 44, 460, 286, 120, 130, 134, 698, 832, 580, 612, and 48, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 859, 801, 239, 273, 557, 23, 1119, 43, 459, 285, 119, 129, 133, 697, 831, 579, 611, 47, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more polynucleotides of: (a) or (b).
2. The transgenic plant of claim 1 wherein the transgenic plant possesses an altered trait as compared to another plant, or the transgenic plant exhibits an altered phenotype as compared to another plant, or the transgenic plant expresses an altered level of one or more genes associated with a plant trait as compared to another plant, wherein the other plant does not comprise the recombinant polynucleotide.
3. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
4. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics,

apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.

5. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
6. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:860.
7. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:802.
8. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:240.
9. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:274.
10. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:558.
11. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:24.
12. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:1120.
13. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:44.

14. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:460.
15. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:286.
16. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:120.
17. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:130.
18. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:134.
19. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:698.
20. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
21. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:580.
22. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:612.
23. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:48.
24. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:859.

25. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:801.
26. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
27. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
28. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
29. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:23.
30. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:1119.
31. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:43.
32. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:459.
33. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
34. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:119.
35. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:129.

36. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:133.
37. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
38. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
39. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:579.
40. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:611.
41. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:47.
42. The transgenic plant of claim 1, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to said nucleotide sequence.
44. The transgenic plant of claim 1, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, mint and other labiates, rosaceous fruits, and vegetable brassicas.
44. The transgenic plant of claim 1 wherein the encoded polypeptide is expressed and regulates transcription of a gene.

45. A method of using the transgenic plant of claim 1 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.
46. An isolated or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:
- (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID NOs: 240, 274, 558, 286, 698, and 832, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 239, 273, 557, 285, 697, 831, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a) or (b).
47. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:240.
48. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:274.
49. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:558.
50. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:286.
51. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:698.

52. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
53. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
54. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
55. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
56. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
57. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
58. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
59. The isolated or recombinant polynucleotide of claim 46, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence.
60. The isolated or recombinant polynucleotide of claim 46 wherein the encoded polypeptide is expressed and regulates transcription of a gene.
61. A vector comprising the isolated or recombinant polynucleotide of claim 46.
62. A host cell comprising the vector of claim 61.

63. A method of using the isolated or recombinant polynucleotide of claim 46 for producing a plant having a modified trait, the method comprising selecting a polynucleotide that encodes a polypeptide, inserting the polynucleotide into an expression vector, introducing the vector into a plant or a cell of a plant to overexpress the polypeptide, thereby producing a modified plant, and selecting a modified plant for a modified trait.
64. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
65. The method of claim 63 wherein the plant possesses a modified as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.
66. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
67. A modified plant produced by the method of claim 63.
68. A method of using the plant of claim 67 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

69. The plant produced by the method of claim 68.

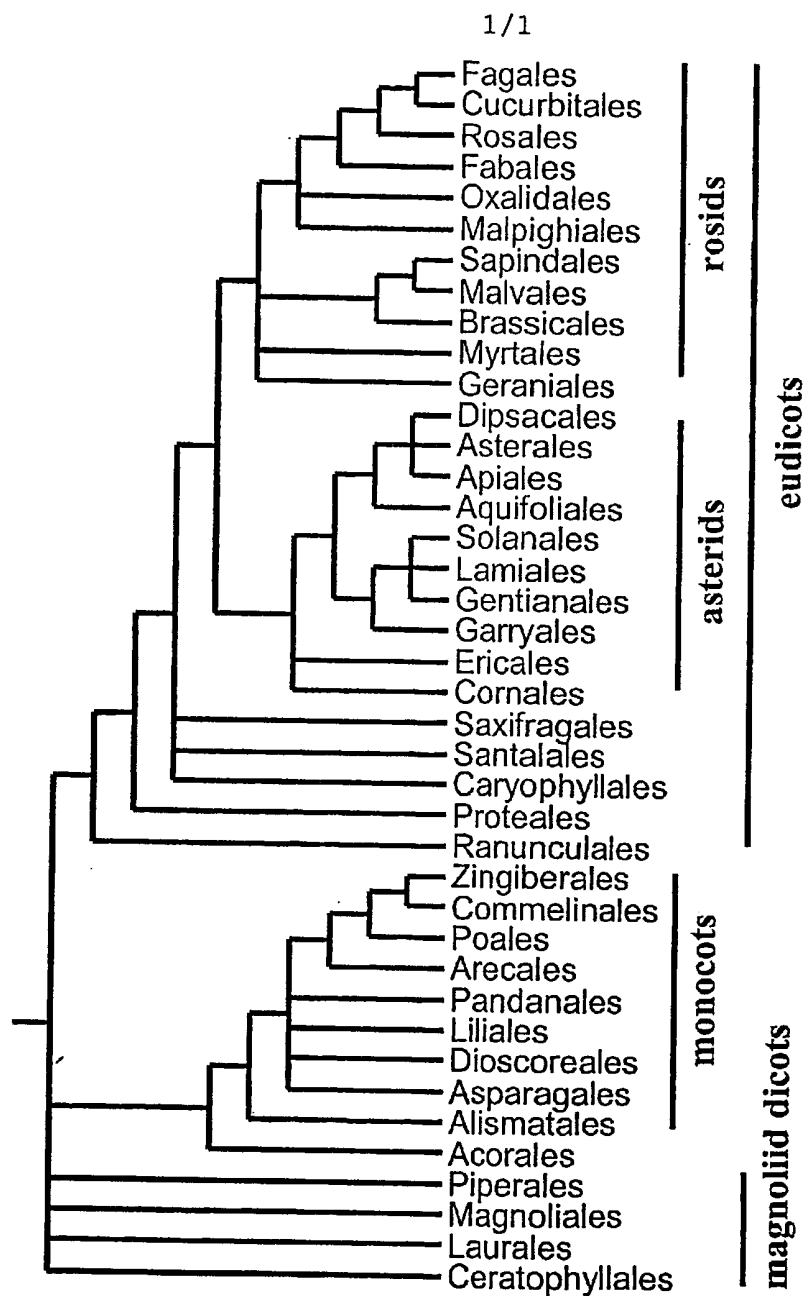


Figure 1

SEQUENCE LISTING

<110> Mendel Biotechnology, Inc.

Ratcliffe, Oliver
Riechmann, Jose Luis
Adam, Luc J.
Dubell, Arnold T.
Heard, Jacqueline E.
Pilgrim, Marsha L.
Jiang, Cai-Zhong
Reuber, T. Lynne
Creelman, Robert A.
Pineda, Omaira
Yu, Guo-Liang
Broun, Pierre E.

<120> YIELD-RELATED POLYNUCLEOTIDES AND
POLYPEPTIDES IN PLANTS

<130> 514442002041

<150> 60/310,847

<151> 2001-08-09

<150> 60/336,049

<151> 2001-11-19

<150> 60/338,692

<151> 2001-12-11

<150> 10/171,468

<151> 2002-06-14

>G1275 (58..579)

CCAAGAAAAGGGGAAGATCACGCATTCTTATAGGCGTAATTCGTAAATAGTGGTGACTATG
AATGATGCAGACACAACTTGGGGAGTAGTTTCAGCGATGATACTCACTCTGTGTTCGAG
TTTCCGGAGCTAGACTTGTCTAGATGAATGGATGGATGATGATCTTGTGTCTGCGGTTTCC
GGGATGAATCAGTCTTATGGTTATCAGACTAGTGATGTTGCTGGTGCTTTATTCTCAGGT
TCTTCTAGCTGTTTCAGTCATCTGAATCTCCAAGTACCAAACTTATGTTGCTGCTACA
GCCACTGCTTCTGCGGACAACCAAAACAAGAAAGAAAAGAAAAAATTAAAGGGAGAGTT
GCGTTCAAGACACGGTCCGAGGTGGAAGTGCTTGACGACGGGTTCAAGTGGAGAAAGTAT
GGGAAGAAGATGGTGAAGAACAGCCACATCCAAGAACTACTACAAATGTTCAAGTTGAT
GGCTGTCCCGTGAAGAAAAGGGTTGAACGAGACAGAGATGATCCGAGCTTTGTGATAACA
ACTTACGAGGGTTCCCAATCACTCAAGCATGAACTAAGACTCGAACTAAGGCTCAAGG
CGACCATGCTATATTCAGCACATCTTATTTTCTATGGTTACGAACGATACTTAAACTGC
TTCTAGTTCTTTATATCCATTGTAAACTGGTTGCAGGTTCAAAATTTTGAGAGGTTTAT
GACATTCTAAATCTGTAGTACTTATATA

>G1275 Amino Acid Sequence (domain in AA coordinates: 113-169)
MNDADTNLGSSFDDTHSVFEFPELDLSDEWMDDLVSAVSGMNQSYGYQTSVAGALFS
GSSSCFSPHESPSTKTYVAATATASADNQNKKEKKIKGRVAFKTRSEVEVLDDGFKWRK
YGKMKVKNSPHPRNYKCSVDGCPVKRVERDRDDPSFVITTYEGSHNHSSMN*

>G1411 (110..856)

TAAAGAAAACCTGAACAACCCCTAAAGTACTGTATAAATCCTATATCAAATTTTTTTTTTG
GAAGAAAAGGCTATATTTAAAGAAAATCAAGCAAAAGTAGATCCTCGGATGTATGGGAA
GAGGCCTTTTGGAGGTGATGAATCTGAAGAAAGGGAAGAAGATGAGAACTTGTTCCTGGT
CTTCTCGGCCCGATCTCAACACGACATGCGTGTTATGGTCTCGGCCTTGACTCAAGTAAT

TCGACTCCTTAATTGCATCACCAACCTAACCCCTCATCATATGATCACCTTCTTTTTTT
CCTTTAATATATATATATGCTCACACACACATATATATATACATATAAGCATCGCCTC
AAGCATTAAAATTTTTACGAACCAAACAAAATAATTATGAATAATTATAATGAACC
CATCTCTCTTCCAAAATTACACTTGGAAACAACATCATCAACAGCAGCAACAACAACA
AGAATGATGATCATCATCATCAACATAATAATGATCCAATCGGTATGGCCATGGACCAGT
ACACAGAGTCCCATATCTTCAATCCTTCTCTTCTTCTCATTTCCCTCCTCTCTTCTT
CCCTCAACAACCCACTCTTCTCTCCGGAGATCAAGAAGACGACGAAGACGAAGAAGAAC
CTCTAGAGGAACTCGGTGCTTATGAAGAAATGATGTACAAGATCGCAGCCATGCAATCGG
TTGACATCGACCCAGCAACCGTCAAGAAACCCAAACCGTAAACGTTAGGAGTCTCCGAC
ACCTCAGAGTGTGGCGGCTAGACATCGCCGTGAGAGAATCAGTGAGAGATCAGAATTC
TTCAGAGACTCGTGCCAGGTGGCACTAAAATGGATACGGCTTCAATGCTCGATGAAGCTA

TACGCTATGTCAAGTTCTTGAAACGGCAGATCCGGCTACTCAATAATAATACCGGATATA
CTCCTCCGCCGCCGCAAGATCAAGCTTCTCAGGCGGTGACGACGTCATGGGTTTCACCGC
CACCACCGCTCAAGTTTCGGCCGTGGGGGAAGAGGAGTAGGAGAATTAATCTAGACAAGAT
GACATTTCCATTAGTAGTAATAAATTATGCTATAATGTGTGAGTAATGGTGCAATTATG
GA

>G1499 Amino Acid Sequence (domain in AA coordinates: 118-181)
MNNYNMNP SLFQNYTWNNIINSSNNNNKNDH HHQHNDPIGMAMDQYTLHIFNPFSSS
HFPPLSSSLTTTLLSGDQEDDEDEEEPLEELGAMKEMMYKIAAMQSVDIDPATVKKPKR
RNVRISSDPQSVAARHRRERISERIRILQRLVPGGTKMDTASMLDEAIRYVKFLKQIRL
LNNNTGYTPPPPDQASQAVTTSWVSPPPPSFGRGGRGVGELI*

>G1543 (1..828)
ATGATAAACTACTATTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAAATATAAAGGCATCGTCACGCTTCAAGTT
TGTCTCTTTTATATTAACTGAGAGTTTTCTCTCAAACCTTACCTTTTCTTCTCGATC
CTAGCTCTTAAGAACCCTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGGAAGAAC
TCTTCAAACCTTGGATCTTACTATCTCCGTTCCAGGCTTCTCTTCATCCCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAAGTCACGATGATGGCTCTGCTCCTCCGCGAAAGAAA
CTCCGTCTAACCAGAGAACAGTCACGCTTCTTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAACAAAAGGAAGTACTTGCCAAGCATTTGATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAACCGTAGAGCAAGGAGCAAATTGAAGCAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTTCAATTAACGGAAGAAAACACAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTCTGCCTCGAGCCTT
ACTATGTGTCTCGCTGCGAGCGAGTTACCCCTGCCGCGAGCCCTTCGAGGGCGGTGGTG
CCGCTTCCGGCTAAGAAAACGTTTCCGCCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)
MIKLLFTYICTYTYKLYALYHMDYACVMYKYKGI VTLQVCLFYIKLRVFLSNFTFSSSI
LALKNPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDEGSGGGRDQLRLDMNRLPSS
EDGDDEEFSHDDGSAPPRKKRLRTREQSRLLDSFRQNH TLNPKQKEVLAKHLMRLPRQI
EVWFQNRARRSKLKQTEMECEYLRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKKTFFPQERDR*

>G1635 (1..1164)
ATGGCGTCGTCTCCGTTGACTGCAAATGTTCAAGGTACCAACGCTTCTTTGAGGAATAGA
GATGAAGAACTGCAGACAAGCAGATACAATTCAATGACCAAAGTTTGGGGGAAATGAC
TATGCACCCAAGGTACGGAAGCCATACAGGATAACAAAAGAGAGAGAGATGGACAGAT
GAAGAGCACAGAAGTTTGTGTAAGCCTTGAAATTATACGGGCGAGCTTGAGAGACGAATA
GAAGAACATGTGGGCTCAAAGACCGCAGTTTCAAGATTGCAAGCCATGCTCAGAAGTTTTC
TCTAAGGTTGCTCGAGAAGCACTGGAGGTGATGGGAGCTCAGTAGAGCCGATTGTAATA
CCTCCTCCTCGTCCCAAGAGAAAGCCAGCGCATCCGTACCCCTCGTAAGTTTGGGAACGAG
GCAGATCAAACAAGTAGATCGGTTTCTCCCTCAGAACGTGATACTCAATCTCAACCTCT
GTGTTGTCCACTGTGATCAGAAGCATTGTGTTCCCTTGATTGAGTTTACCCAATCGA
AGCTTGTCCCCAGTTTCTTCTGCATCACCACCAGCTGCTCTTACAACCACTGCAAATGCA
CCTGAAGAGCTTGAGACTCTGAAGCTGGAGTTGTTTCTAGTGAGAGACTCTTAAACAGG
GAGAGCTCGATCAAGGAACCAACGAAGCAAAGTCTTAAACTCTTTGGGAAGACAGTTTG
GTATCTGATTCAGGCATGTCTCTTCTCTAACAACCTTCAACATATTGTAAATCCCAATT
CAGCCATTACCACGGAACCTCTCATCATCAAGACACTACCCATAATAAGAACTCACAA
GAAGAACCTTTGAGCTGCTGGATACAAGTCCCTCTTAAGCAAGAAGATGTGGAAAATAGA
TGTTTGGATTTCAGGAAAGGCTGTCCAAAACGAAGGATCATCGACTGGATCAAACACTGGT
TCGGTGGATGATACGGGACACACGGAAGACACAGAACCCGAAACAATGCTATGTCAA
TGGGAGTTTAAACCAAGTGAGAGGTCTGCATTTTCTGAGCTCAGAAGAACAACCTCCGAG
TCAATTCAAGAGGATTTGGTCCATACAAGAAGAGAAAGATGGTAACAGAAGAAGAAGAG
CATGAGATTCACTCTCACTTATAA

>G1635 Amino Acid Sequence (domain in AA coordinates: 44-104)
MASSPLTANVQGTNASLRNRDEETADKQIQFNDQSFGGNDYAPKVRKPYTITKERERWTD
EEHKFVEALKLYGRAWRRIEEHVSGKTAVQIRSHAQKFFSKVAREATGGDGSSVEPIVI
PPRPKRKPAHPYPRKFGNEADQTSRSVSPSERDTQSPTS SVLSTVGSEALCSLDSSSPNR
SLSPVSSASPPAALTTTANAPEELET LKLELFP SERLLNRESSIKEPTKQSLKLFKTVL

VSDSGMSSSLTTSTYCKSPIQPLPRKLSSSKTLPIIRNSQEELLSCWIVPLKQEDVENR
CLDSGKAVQNEGSSTGSGVDDTGHTEKTEPETMLCQWEFKPSERSAFSELRRNTNSE
SNSRFGPGYPYKKRKMVTEEEHEIHLHL*

>G1794 (160..1335)

TCTTTCTTTCTTCTCTTTGCTCTGTTTCTGTTTCTCTCTCTCTCTCTACAGAGTT
TTCTTTCCCTCGAAGAAAAAGAATATTTTAAATTTAATTTTCTCTGCGTTTATAAGCTT
TAAGTTTCAGAGGAGGAGGATTTAGAAGGAGGGTTTGTATGTGTGCTTAAAAGTGGCA
AATCAGGAAGATAACGTTGGCAAAAAAGCCGAGTCTATTAGAGACGATGATCATCGGACG
TTATCTGAAATCGATCAATGGCTTTACTTATTTCGAGCCGAGACGACCACCACGTCAT
AGCTTCCCTACGCAGCAGCCGCCTCCATCGTCGTCCTCATCTCTTATCTCAGGTTTC
AGTAGAGAGATGGAGATGTCTGCTATTGTCTCTGCTTTGACTCACGTTGTTGCTGGAAAT
GTTCTCAGCATCAACAAGGAGGAGAGAGGTGGAGGAAGGTGGCGCCAAAGCGGTTAAG
TCTTCTCGGGGCGAAGGAGGAGAGGTGGAGGAAGGTGGCGCCAAAGCGGTTAAG
GCAGCTAATACTTTGACGGTTGATCAATATTTCTCCGGTGGTAGCTCTACTTCTAAAGTG
AGAGAAGCTTCGAGTAACATGTGAGTCCGGGCCCAACATACGAGTATACAACTACGGCA
ACTGCTAGTAGCGAAACGTCGTCGTTTACTGGGGACCAACCTCGGCGAAGATACAGAGGA
GTTAGACAAAGACCATGGGGAAAGTGGGCGGCTGAGATTGAGATCCATTTAAAGCAGCT
AGAGTTTGGCTCGGTACGTTGACAATGCTGAATCAGCAGCAAGAGCTTACGACGAAGCT
GCACTTCGGTTTAGAGGCAACAAAGCCAAACTCAACTTCCCTGAAAACGTCAAACTCGTT
AGACCTGCTTCAACCGAAGCACAACCTGTGCACCAACCGCTGCTCAAAGACCGACCCAG
TCAAGGAACCTCGGGTTCAACGACTACCTTTTGGCCATAAGACCTGCTTCAATCAAAGC
GTTCAATTCGAGCCGTTGATGCAATCATACAATCTTTGGATTATACGATCAAATGTCGTTT
CAACAACAGTTTCAGCAACATCATCAACAATCTTTGGATTATACGATCAAATGTCGTTT
CCGTTGCGTTTCGGTCACACTGGAGGTTCAATGATGCAATCTACGTCGTCATCATCATCT
CATTCTCGTCTCTGTTTTCCCGGCTGCTGTTTCCGCGCCACCAGAATCAGCTAGCGAA
ACCGGTTATCTCCAGGATATACAATGGCCATCAGACAGACTAGTAATAACTACAATAAT
AGTCCATCCTCCTGATGACTTGCTTCATTTTATTTGTTTCACTATAGAGTAATAGAAAAC
AGGAAAATGATTATATGTTATAGAGTTATTTTCCAAATATTATAGGGTTTAGGTTGTTT
GTATTGTTCTGCTTTTCATCCTCTCATGCTTTTTTCTTAATTTATTATTTTTGCATTA
TAATTTTCGTTTCATTGTAACAAACATTAAGAACACATGGAGAAAGGAAAAAAGAG
AG

>G1794 Amino Acid Sequence (domain in AA coordinates: TBD)

MCVLKVANQEDNVGKKAESIRDDHRTLSEIDQWLYLFAAEDDHHRHSFPTQQPPSSSS
SSLISGFSREMESAIVSALTHVAGNVPQHQQGGEGSGEGTSNSSSSSGQRRREVEE
GGAKAVKAANTLTVDQYFSGSSSTSKVREASSNMSGPGPTYEYTTTATASSETSSFSGDQ
PRRRYRGVRQRPWGKWAEEIRDPFKAARVWLGTFDNAESAARAYDEAALFRGNKAKLNF
PENVKLVRPASTEAPVHQTAQRPTQSRNSGTTTTLLPIRPASNQSVHSQPLMQSYNLS
YSEMARQQQFQQHQQSLDLVDQMSFPLRFHTGGSMQSTSSSSSHSRPLFSPAAPQP
PPESASETGYLQDIQWPSDKTSNNYNNSSPS*

>G1839 (38..592)

ATCACAGTTATGTTTCCATTTCATGCGCTATAAAAACCATGCTCACTCCCTTTTGTTCTTC
ACACCATTTGCAGGAAAAAATGAATAGTTGTCAGTCTAATCCCACCAAAATGGATAATTC
AGAAAATGTTCTATTTAATGATCAAAACGAAAATTTACACTTGTGTCACCACACCCTTC
TTCTTCGTACTTGACAAGAGATCAAGAGCAGGATCATGGTCTCTGCTCTGCGACAAGT
GATATCTAACTCCGGAGCTGACGACGCGTCATCATCAAACTTGATCATCACAAGCGTTCC
GCCTCCAGACGCTGGCCCTTGTCTCTCTGTTGGCGTCCGCGTTGCTACGGCTGCACATT
ACAACGGCCCGACCGAGAGGTAAAGAAGGAGAAGAAATACAAGGAGTAAGGAAAAAACC
ATCGGGTAAATGGGEGGCGGAGATATGGGATCCGAGATCGAAATCAAGGAGGTGGCTTGG
AACGTTTCTTACGGCGGAGATGGCGGCACAATCTTACAATGATGCGGCGGCTGAGTATCG
AGCAAGACGTGGTAAAACAAACGGAGAAGGAATTAAACGGCGGTGGAGATGACTGAGAAG
GACATGGTCCGGTATCATACACGGCGAGGTGGAAATGTTATTTACTATTGAAAATAA
ATTATTTATTATAGAGGGAGATATTACTCTTTACGCTTTTCATTAGATTTATTTTTATAA
GTTTTAAAGTATTTTATTGTTATAAAAAA

>G1839 Amino Acid Sequence (domain in AA coordinates: TBD)

MLTPFCSSHLQEKMNCSQSNPTKMDNSENVLFDQENFTLVAPHPSSSYLTRDQEHEI
MVSALRQVISNSGADDASSNLIITSVPPPDAGPCPLCGVAGCYGCTLQRPHEVKKKK
YKGVRRKPSGKWAEEIWDPRSKSRRWLGTFLTAEMAAQSYNDAAEYRARRGKTNGEIK

RRWR*

>G2108 (35..694)

GAGAGAGAAACATTGATCTCTGAATATTGTGAACATGTTGAAATCAAGTAACAAGAGAAA
 AAGCAAAGAAGAGAAGAAGTTACAAGAAGGGAAGTACCTTGGAGTGAGGAGACGTCCATG
 GGAAGATATGCAGCTGAAATCAGAAACCCCTTTTACTAAAGAAAGACATTGGCTTGAAC
 GTTTGATACAGCCGAAGAAGCTGCTTTTGCATATGACGTTGCTGCTCGATCCATCAGCGG
 CTCTCTAGCTACAACAACTTCTTCTACACTGAAAACACCTCTTTAGAAAGACATCCACA
 ACAGTCTTTGGAGCCTCATATGACTTGGGGATCTTCTAGTCTCTGTCTTCTTCAAGATCA
 GCCTTTTGAACAAACCATTTTGTGCTGATCCTATCTCTTCTTCTTTTCTCAAAAACA
 AGAGTCTTCTACCAATCTCACTAACACTTCTCACATTGTTATAATGATGGTGATCATGT
 TGGCCAAAGCAAAGAGATTTCTTTACCTAATGATATGTCAAACAGTTTATTCCGTCATCA
 GGACAAAGTCGGTGAACATGACAAATGCAGACCATATGAAGTTTGGCTCAGTTCTCAGCGA
 CGAACCTCTCTGCTTTGAGTATGACTACATTGGGAATTATCTTCAGAGTTTCTCAAAGA
 TGTCAACGACGATGCTCCACAGTTTCTTATGTGAGCTTGATTACCGATCCTTCAATTTA
 TG

>G2108 Amino Acid Sequence (domain in AA coordinates: 18-85)

MLKSSNKRKSKKEKKLQEGKYLGVRRRPWGRYAAEIRNPFKERHWLGTFTDTEEAFAFAY
 DVAARSISGLATTNFFYTENTSLEHRPQQSLEPHMTWGSSSLCLLDQPFENNHFVADP
 ISSFSQKQESSNTLNTFNSHCYNDGDHVGQSKEISLPNDMSNSLFGHQDKVGEHDNADH
 MKFGSVLSDEPLCFEYDIIGNYLQSFLLKDVNDAPQFLM*

>G2291 (27..797)

GCTTTCTCACCTTTTATAAAATAGAAAATGGAAAACAGCTACACCGTTGATGGTCACCGTC
 TTCAATATTCGGTTCCGTTAAGCTCCATGTCATGAAACAGTCAAACTCCGAAACTTACG
 GATTATCCAAAGAGTCGCCGTTGGTCTGTCATGCCCTTGTTCGAAACCAACACTACTTCAT
 TCGATATCTCTTCTCTTTTCTCGTTTAAACCCAAAACAGAACCCGAAAACACGCATCGTG
 TCATGGACGATTCCATCGCCGCCGTCGTGGGCGAAAACGTTCTTTTCCGTGATAAAAACA
 AAGTCTCTGATCACTTGACCAAAGAGGTGGTGTGAAGCGGGGGCGGAAGATGCCGAGA
 AGACCGGAGGATTTCATGGGAGTGAGAAAACGGCCGTGGGGGAGATGGTCGGCGGAGATAA
 GAGACAGGATAGGGCGGTGCAGACACTGGTTAGGAACGTTTCGACACGGCGGAAGAGGCAG
 CGCGTGGTATGACGCGGCGGCGAGGAGGCTTAGAGGGACCAAAGCCAAGACCAATTTCTG
 TGATTCTCCGCTTTTCCCAAGGAAATAGCTCAGGCTCAGGAGGATAATAGGATGAGGC
 AGAAGCAGAAGAAGAAGAAGAAAAAGTGAGTGTGAGGAAGTGTGTTAAAGTCACAT
 CGGTTGCACAGTTGTTTCGATGATGCCAATTTTATAAATCTTCTAGTATTAAAGGAAATG
 TGATTAGTTCTATTGATAATCTTGAAAAATGGGTCTAGAGCTTGATTGAGTTAGGGT
 TGTTGTCTAGGAAGTGATAAAGCACTCGTAGTTAAGTAGTTGTAGTT

>G2291 Amino Acid Sequence (domain in AA coordinates: TBD)

MENSYTVDGHRLLQYSVPLSSMHETSQNSETYGLSKESPLVCMPLFETNTTSFDISSLSF
 NPKPEPENTHRVMDSDIAAVGVENVLFGDKNKVSDHLTKEGGVKRGKMPQKTGGFMGVR
 KRPWGRWSAEIRDRIGRHNLGTFTDTEEAARAYDAAARRLRGTAKTNFVIPPLFPKE
 IAQAQEDNRMROKQKQKKKKKVSVRKCVKVTVAQLFDDANFINSSSIKGNVISSIDNLE
 KMGLELDLSLGLLSRK*

>G2452 (1..804)

ATGTCATCGTCGACGATGTACAGAGGAGTTAATATGTTTTTACC CGGCAAACACAAACTGG
 ATTTTTCAAGAAGTCAGAGAAGCCACGTGGACGGCGGAGGAGAACAACGGTTCGAGAAA
 GCTCTCGCTTATCTGGACGACAAAGACAATCTTGAGAGCTGGTCCAAGATCGCAGATTTG
 ATTCCCGGCAAAACAGTAGCTGACGTCATTAAACGATACAAGGAGCTAGAGGATGATGTC
 AGCGACATCGAAGCCGGACTTATCCCCATTCCGGGATACGGCGGCGACGCTCCTCCGCT
 GCAAAACAGTGACTATTCTTTGGTCTAGAAAACCTCAGCTACGGTTATGATTACGTCGTT
 GGAGGAAAGAGGAGTTTCGCCGCGGATGACTGATTGTTTAGGTCTCCGATGCCGGAAG
 GAGAGGAAGAAAGGAGTTCCGTGGACCGAGGACGAACACCTACGATTCTGATGGGTTTG
 AAGAAATATGGAAGAGGAGATTGGAGAAACATAGCAAAAAGCTTTGTGACGACTCGAACG
 CCGACGCAAGTCGCTTCACACGCTCAGAAATATTTCTTCGACAACTCACAGATGGTAAA
 GACAAAAGACGATCAAGTATTCACGATATCACCCTGTAAACATCCCTGACGCAGACGCA
 TCCGCAACCGCCACGACCGCTGACGTAGCACTCTCTCTACTCCAGCCAATCTTTTGAC
 GTTTTCTTCAGCCAAATCCTCATTACAGTTTCGCGTCTGCGTCTGAGCTATTAT
 AATGCGTTTCCGCAGTGGAGTTAA

>G2452 Amino Acid Sequence (conserved domain in AA coordinates: 27-213)

MSSSTMYRGVNMFS PANTNWIFQEVREATWTAENKRFEKALAYLDDKDNLESWSKIADL
IPGKTVDVADVIKRYKELEDDVSDIEAGLIPIPGYGGDASSAANSYDFFGLENSSYGYDYVV
GGKRSSPAMTDCFRSPMPKEKRRKGVPTWTEDEHLRFLMGLKKYKGDWRNIAKSFVTRRT
PTQVASHAQKYFLRQLTDGKDKRRSSIHDITTVNIPDADASATATTADVALSPTPANSFD
VFLQPNPHYSFASASASSYNAFPQWS*

>G2509 (143..934)

ATATATTCCTCTTTCATTCTCCTTCTTCGTCTTTTCTTTGTTTCTCATATTCAAGACAT
CCTCAATTCCAAATCTTAAACCCTAAATTACAGACACAATCGAGATCACCTGAAAAAG
AGGTTTAAAGATTTTAGCAAAGATGGCGAATTCAGGAAATTATGGAAAGAGGCCCTTTCG
AGGCGATGAATCGGATGAAAAGAAAGAAGCCGATGATGATGAGAACATATTCCCTTCTT
CTCTGCCCGATCCCAATATGACATGCGTGCCATGGTCTCAGCCTTGACTCAAGTCATTGG
AAACCAAGCAGCTCTCATGATAATAACCAACATCAACCTGTTGTGTATAATCAACAAGA
TCCTAACCACCGGCTCCTCCAACCTCAAGATCAAGGGCTATTGAGGAAGAGGCACTATAG
AGGGGTAAGACAACGACCATGGGGAAAAGTGGGCAGCTGAAATTCGGGATCCGCAAAAGGC
AGCACGGGTGTGGCTCGGGACATTTGAGACTGCTGAAGCTGCGGCTTTAGCTTATGATAA
CGCAGCTCTTAAGTTCAAAGGAAGCAAAGCCAACTCAATTTCCCTGAGAGAGCTCAACT
AGCAAGTAACACTAGTACAACCTACCGGTCCACCAAACTATTATTCTTCTAATAATCAAAT
TTACTACTCAAATCCCGAGACTAATCCGCAAACCATACTTATTTTAAACCAATACTACTA
TAACCAATATCTTCATCAAGGGGGGAATAGTAACGATGCATTAAGTTATAGCTTGCCCGG
TGGAGAAACCGGAGGCTCAATGTATAATCATCAGACGTTATCTACTACAAATTTCTTCATC
TTCTGGTGGATCTTCAAGGCAACAAGATGATGAACAAGATTACGCCAGATATTTGCGTTT
TGGGGATTCTTCACCTCCTAATTCTGGTTTTTGGAGATCTTCAATAAACTGATAATAAAGG
ATTTGGGTCACTTGTATGAGGGGATCATATGTTTTCTAA

>G2509 Amino Acid Sequence (domain in aa coordinates: 89-156)

MANSGNYGKRPFGRDESDEKKEADDDENIFPFFSARSQYDMRAMVSALTQVIGNQSSSHD
NNQHQPVVYNQODPNPPAPPTQDQGLLRKRHYRGVRQRPWGKWAAEIRDPPQKAARVWLGT
FETAEEAALAYDNAALKFKGSKAKLNFPERAQLASNTSTTTGPPNYYSNNQIYYSNPQT
NPQTIPIYFNQYYNQYLHQGGNSNDALSYSLAGGETGGSMYNHQTLSSTNSSSSGGSSRO
QDDEQDYARYLRFGDSSPPNSGF*

>G390 (1..2526)

ATGATGGCTCATCACTCCATGGACGATAGAGACTCTCCTGATAAAGGATTTGATTCCGGC
AAGTACGTTAGATACACGCCGGAACAAGTTGAAGCTCTTGAGAGAGTTTATGCTGAGTGT
CCTAAACCTAGCTCTCTGAGAAGACAACAGCTTATTTCGTGAATGTCCCATTTCTCTGTAAC
ATCGAGCTCGACAGATCAAAGTTTGGTTCCAGAATCGCAGATGTGAGAGAAAGCAGAGG
AAAGAGTCAGCTCGTCTTCAGACAGTGAACAGGAAGCTGAGTGCTATGAACAAGCTTTTG
ATGGAAGAGAATGATCGTTTGCAGAAGCAAGTCTCCAACCTGGTTTTATGAGAATGGATTCT
ATGAAACATCGAATCCACACTGCTTCTGGGACGACCACAGACAACAGCTGTGAGTCTGTG
GTCGTGAGTGGTCAGCAACGTCAGCAGCAAAACCAACACATCAGCATCCTCAGCGTGAT
GTTAACAACCCAGCTAATCTTCTCTCGATTGCGGAGGAGACCTTGCGGAGTTCTCTTTCG
AAGGCTACAGGAAGTGTGTCGACTGGGTCCAGATGATTGGGATGAAGCCTGGTCCGGAT
TCTATTGGTATCGTAGCTGTTTCACGCAACTGCAGTGGAATAGCAGCACGTGCCTGTGGC
CTCGTGAGTTTAGAACCCATGAAGGTGCTGAAATCCTCAAAGATCGTCCATCTTGGTTC
CGTGACTGTGCGATGTGTCGAGACTCTGAATGTTATACCCACTGGAAATGGTGGTACTATC
GAGCTTGTCAACACTCAGATTTATGCTCCTACACATTAGCAGCAGCTCGTGACTTTTGG
ACGCTGAGATATAGTACAAGTCTAGAAGATGGAAGCTATGTGGTCTGTGAGAGATCACTC
ACTTCTGCAACTGGTGGCCCCAATGGTCCACTTTCTTCAAGCTTCGTGAGAGCCAAAATG
CTGTCAAGCGGGTTTCTTATCCGTCCTTGTGATGGTGGTGGTCCATTATTACATCGTT
GATCATGTGGACTTGGATGTCTCAAGTGTTCTGAAAGTCTCAGGCCTCTTTATGAGTCT
TCCAAAATCCTTGCTCAAAAAATGACTGTGCTGCTCTGAGACATGTGCGCCAAATTGCT
CAAGAGACTAGTGGAGAAGTCCAGTATAGTGGTGGACGCCAGCCTGCAGTTTTAAGGACT
TTCAGCCAGAGACTCTGCCGGGTTTCAATGATGCTGTAAATGGTTTTGTGCGATGATGGA
TGGTCTCCAATGAGTAGTGATGGAGGAGAGGATATTACGATCATGATTAACCTTCTCTCT
GCTAAATTTGCTGGCTCCCAATACGGTAGCTCATTTCTTCCAAGTTTTTGAAGTGGTGTCT
CTCTGTGCCAAAGCTTCTATGCTGTTGCAGAAATGTTCCACCCCTTGTATTGATTGGTTCT
CTGAGAGAACACCGAGCTGAATGGGCAGACTATGGTGTGATGCCTATTCTGCTGCATCT
CTCAGAGCAACTCCATATGCTGTTCCATGCGTCAGAACCGTGGGTTCCCGAGTAACCAA
GTCATTTCTTCTCTCGCACAGACTCGAACATGAAGAGTTTCTCGAAGTGGTTAGACTT

GGAGGTCATGCTTACTCACCTGAAGACATGGGCTTATCCCGGGATATGTATTTACTGCAG
 CTTTGTAGCGGCGTTGATGAAAATGTGGTTGGAGGTTGTGCTCAGCTTGTCTTTGCCCA
 ATCGATGAATCATTGCTGATGATGCACCTTGCTTCCTTCTGGTTTCCGTGTCATACCA
 CTCGACCAAAAAACAAATCCGAATGATCATCAATCTGCAAGTCGAACACGGGATCTAGCA
 TCGTCCCTAGATGGTTCCACCAAAACCGATTCCGGAACAACTCTAGATTGGTCTTAACA
 ATAGCCTTCCAGTTACGTTTGATAAACCATTCCAGAGACAATGTTGCTACAATGGCGAGA
 CAGTATGTGAGGAACGTTGTTGGTTTCGATTGAGAGAGTGGCTCTAGCCATTACGCCCTCGT
 CCTGGCTCAATGCAACTTCCCCTTCCCCTGAAGCTCTCACTCTTGTCCGTTGGATCACC
 CGTAGTTACAGTATTCATACAGGTGCAGATCTGTTTGGAGCTGATTCTCAGTCCTGTGGA
 GGAGACACATTGCTTAAGCAACTCTGGGACCATAGTGATGCCATATTGTGCTGCTCCCTG
 AAACTAATGCCTCACCGGTATTACATTGCAAACCAAGCTGGTTTAGACATGCTTGAA
 ACTACACTTGTGGCACTTACAGGATATAATGCTCGACAAAACACTTGATGACTCTGGTCGT
 AGAGCTCTTTGCTCCGAGTTCGCCAAGATCATGCAGCAGGGATATGCGAATCTTCCGGCA
 GGAATATGTGTGTCGAGCATGGGCAGACCGGTTTCGTATGAGCAAGCGACGGTGTGGAAA
 GTTGTGATGACAACGAATCAAACCACTGCTTGGCTTTTACCCTCGTTAGTTGGTCGTTT
 GTTTGA

>G390 Amino Acid Sequence (domain in AA coordinates: 18-81)
 MMAHHSMDRRDSPDKGFDGSGKYVRYTPEQVEALERVYAECPKPSSLRRQLIRECPILCN
 IEPRQIKVWFQNRRCREKQKESARLQTVNRKLSAMNKLMEENDRLQKQVSNLVYENG
 MKHRIHTASGTTTNDSCSESVVVGQQRQQNPTHQHPQRDVNNPANLLSIAEETLAEFLC
 KATGTAVDQWQVIMGKPGPDSIGIVAVSRNCSGIAARACGLVSLPEPMKVAEILKDRPSWF
 RDCRCVETLNVIPITGNGGTIELVNTQIYAPTTLAAARDFWTLRYSTSLDGSYVVCERSL
 TSATGGPNGPLSSSFVRKMLSSGFLIRPCDGGGSIHIVDHDLDVSSVPEVLRPLYES
 SKILAQKMTVAALHRVQIAQETSSEVQYSGGRQPAVLRTFSQRLCRGFNDVNGFVDDG
 WSPMSSDGGEDITIMINSSAKFAGSQYGSFLPSFGSGVLCASMLLQNVPLVLIRF
 LREHRAEWADYGVDAISAASLRATPYAVPCVRTGGFSPSNQVILPLAQTLHEEFLEVRL
 GGHAYSPEDMGLSRDMYLLQLCSGVDENVVGGCAQLVFAPIDESFADDAPLLPSGFRVIP
 LDQKTNPNHDQSASRTRDLASSLDGSTKTDSETNSRLVLTIAFQFTFDNHSRDNVATMAR
 QYVRNVVGSIQRVALAITPRPGSMQLPTSPEALTLVRWITRSYSIHTGADLFGADSQSCG
 GDTLLKQLWDHSDAILCCSLKTNASPVFTFANQAGLDMLETTLVALQDIMLDKTLDDSGR
 RALCSEFAKIMQQGYANLPAGICVSSMGRPVSYEQATVWKVVDNENHCLAFTLVSWSF
 V*

>G391 (1..2559)

ATGATGATGGTCCATTTCGATGAGCAGAGATATGATGAACAGAGAGTCGCCGGATAAAGGG
 TTAGATTCCGGCAAGTATGTGAGGTACACGCCGGAGCAAGTGGAAGCTCTCGAGAGAGTT
 TACACTGAGTGTCTTAAGCCAAGTTCTCTAAGAAGACAACAACCTCATACGTGAATGTCCG
 ATTCTCTCTAACATCGAGCCTAAGCAGATCAAAGTTTGGTTTCAGAACCGCAGATGTCTG
 GAGAAGCAGAGGAAAGAGCTGCTCGTCTTCAAACAGTGAACAGAAAACCTCAATGCCATG
 AACAACTCTTGATGGAAGAGAATGATCGTTTGCAGAAGCAAGTTCTAACTTGGTCTAT
 GAGAATGGCCACATGAAACATCAACTTCACACTGCTTCTGGGACGACCACAGACAACAGC
 TGTGAGTCTGTGTCGTGAGTGGTCAGCAACATCAACAGCAAAACCCAAATCCTCAGCAT
 CAGCAACGTGATGCTAACAAACCCAGCAGGACTCCTTTCTATAGCAGAGGAGGCCCTAGCA
 GAGTTCCTTTCCAAGGCTACAGGAAGTCTGTTGACTGGGTTCAGATGATTGGGATGAAG
 CCTGGTCCGGATTCTATTGGCATAGTCGCTATTTTCGCGCAACTGCAGCGGAATTGCAGCA
 CGTGCTGCGGCCCTCGTGAGTTTAGAACCCATGAAGGTTGCTGAAATTCTCAAAGATCGT
 CCATCTTGGCTCCGAGATTGTGCAAGTGTGGATACTCTGAGTGTGATACCTGCTGGAAAC
 GGTGGGACGATCGAGCTTATTTACACGCAGATGTATGCTCCTACGACTTTAGCAGCAGCT
 CGTGACTTTTGGACGCTGAGATATAGCACATGTTTGAAGATGGAAGCTATGTGGTTTGT
 GAAAGGTCGCTTACTTCTGCAACTGGTGGCCCCACTGGGCCACCTTCTTCAAACCTTTGTG
 AGAGCTGAAATGAAACCAAGCGGGTTTCTCATCCGTCCTTGGCATGGTGGTGGTTCCATT
 CTCACATTGTTGATCATGTTGATCTGGATGCTTGGAGTGTCCCTGAAGTCATGAGGCCT
 CTCTATGAATCATCGAAGATTCTTGCTCAGAAAATGACTGTTGCTGCTTTGAGACATGTA
 AGACAAATTGCACAAGAAACAAGTGGAGAAGTTCAGTATGGTGGAGGGCGCCAACCTGCCG
 GTTTTAAGAACCTTCAGTCAAAGACTCTGTGCGGGTTTCAATGATGCTGTTAATGGTTT
 GTGGATGATGGATGGTACCAATGGGTAGCGATGGTGCAGAGGATGTTACTGTAATGATA
 AACTTGTCCCTTGGGAAGTTTGGTGGGTCTCAGTACGGTAATTATTCCTTCCAAGCTTT
 GGTAGTGGCGTGCTTTGTGCCAAGGCATCTATGTTGCTTCAGAACGTTCCACCCGCTGTG

CTGGTTCGATTCCCTTAGAGAACACCGATCTGAATGGGCTGATTATGGCGTGGATGCTTAT
GCTGCTGCATCGCTCAGAGCAAGTCCTTTTGCTGTTCCCTTGCTAGAGCTGGGGGGTTC
CCAAGTAACCAAGTCATTCTTCTCTTGCAGACAGTTGAACATGAAGAGTCACTTGAG
GTGGTTAGACTTGAAGGTACGCTTACTACCCGAAGACATGGGTTTAGCTCGGGATATG
TATTTGCTACAGCTTTGTAGCGGTGTTGATGAAAATGTGGTTGGAGGTTGTGCACAGCTT
GTATTTGCCCCATATCGATGAATCATTTGCTGATGATGCACCTTTGCTTCCCTCCGGTTTC
CGCATCATACCTCTTGAACAGAAATCTACTCCGAACGGTGCATCTGCAAACCGTACCCTG
GATTTAGCCTCAGCTTTAGAAGGATCCACAGTCAAGCTGGTGAAGCCGACCCAAATGGC
TGTAACCTTTAGTCCGCTACTAACCATAGCATTCCAGTTCACATTTGATAACCATTCAAGA
GACAGTGTGCTTCAATGGCACGTACGTACGTGCGAAGCATAGTAGGATCGATTAGAGG
GTTGCTCTAGCCATTGCTCCTCGTCCGCTCCGCTCCAATATCAGTCCAATATCTGTCCCACT
TCCCCTGAAGCTCTCACTCTGGTCCGTTGGATCTCCCGGAGTTACAGCCTTCACACTGGT
GCAGATCTCTTTGGATCTGATTCTCAAACAGTGGTGACACGTTGCTGCATCAACTCTGG
AATCACTCTGATGCAATCTTGTGCTGCTCCCTCAAACAAACGCTTACCGGTTTTCACA
TTCGCAAACCAAACCGGTTTAGACATGCTGGAACGACTCTTGTAGCCCTTCAAGACATA
ATGCTAGACAAGACCTTACGAACCTGGTCGTAAGCTCTTGTCTGAGTTCCCAAG
ATCATGCAACAGGGCTATGCTCATCTGCCGCGAGGATATGTGCGTCAAGCATGGGAAGG
ATGGTATCTTACGAGCAGGCAACGGTGTGGAAGTTCTTGAAGACGATGAATCAAACAC
TGCTTAGCTTTTCATGTTCTGTAATTGGTCTGTTCTGTTGA

>G391 Amino Acid Sequence (domain in AA coordinates: 25-85)
MMVHMSRDMNRESPTDKLDSGKYVRYTPEQVEALERVYTECPKPSSLRRQQLIRECP
ILSNIEPKQIKVWFQNRRCREKQKKEAARLQTVNRKLNAMNKLMEENDRLQKQVSNLVY
ENGHMKHQLHTASGTTTNDSCSVVVSQGHQHQNNPNPQHQQORDANNPAGLLSIAEEALA
EFLSKATGTAVDWVQMIGMKPGPDSIGIVAISRNCSGIAARACGLVSLPEMKVAEILKDR
PSWLRCRSDVTLSPVAGNGGTIELIYTQMYAPTTLAAARDFWTLRYSTCLEGDSYVVC
ERSLTSATGGPTGPPSSNFVRAEMKPSGFLIRPCDGGGSILHIVDHVDLDAWSVPEVMRP
LYESSKILAQKMTVAALRHVRQIAQETSSEVQYGGGRQPAVLRTFSQRLCRGFNDVNGF
VDDGWSMPGSDGAEDVTVMINLSPGKFGGSQYGNFSLPSFGSGVLCAKASMLLQNVPPAV
LVRFLREHRSEWADYGVDAYAAASLRASPFAPPCARAGGFPSNQVILPLAQTVHEESLE
VVRLEGHAYSPEDMGLARDMYLLQLCSGVDENVVGGCAQLVFAPIDESFADDAPLLPSGF
RIIPLEQKSTPNGASANRTLDLASALEGSTRQAGEADPNCGNFRSVLTIAFQFTFDNHSR
DSVASMARQYVRSIVGSIQRVALAIAPRPGSNISPISVPTSPEALTLRVWISRSYSLHTG
ADLFGSDSQTSGDTLLHQLWNHSDAILCCSLKTNASPVFTFANQTGLDMLETTLVALQDI
MLDKTLDEPGRKALCSEFPKIMQQGYAHLPAVCASSMGRMVSYEQATVWKVLEDDSNH
CLAFMFVNWSFV*

>G438 (188..2716)
CGGGGTACCCAAGCCACGACCGTAGAATCTTCTTTTGTCTGAAAAGAATTACAATTTACG
TTTCTCTTACGATACGACGGACTTTCCGAAGAAATTAATTTAAAGAGAAAAGAAGAA
GCCAAAGAAGAAGAAGCTAGAAGAAACAGTAAAGTTTGAGACTTTTGTGAGGGTCG
AGCTAAAATGGAGATGGCGGTGGCTAACCACCGTGAGAGAAGCAGTGACAGTATGAATAG
ACATTTAGATAGTAGCGGTAAGTACGTTAGGTACACAGCTGAGCAAGTCGAGGCTCTTGA
GCGTGTCTACGCTGAGTGTCTTAAGCCTAGCTCTCTCCGTCGACAACAATTGATCCGTGA
ATGTTCCATTTTGGCCAAATATTGAGCCTAAGCAGATCAAAGTCTGGTTTCAGAACCAGCAG
GTGTCGAGATAAGCAGAGGAAAGAGGCGTCGAGGCTCCAGAGCGTAAACCGGAAGCTCTC
TGCGATGAATAAACTGTTGATGGAGGAGAATGATAGGTTGCAGAAGCAGGTTTCTCAGCT
TGTCTGCGAAAATGGATATATGAAACAGCAGCTAACTACTGTTGTTAACGATCCAAGCTG
TGAATCTGTGGTCACAACCTCCTCAGCATTTCGCTTAGAGATGCGAATAGTCTGCTGGATT
GCTCTCAATCGCAGAGGAGACTTTGGCAGAGTTCTATCCAAGGCTACAGGAACGTCTGT
TGATTGGGTTTCAGATGCTGGGATGAAGCCTGGTCCGGATTCCGTTGGCATCTTTGCCAT
TTCGCAAAGATGCAATGGAGTGGCAGCTCGAGCCTGTGGTCTTGTAGCTTAGAACCTAT
GAAGATTGCAGAGATCCTCAAAGATCGGCCATCTTGGTTCCGTACTGTAGGAGCCTTGA
AGTTTTCATATGTTCCCGGCTGGTAATGGTGGCACAATCGAGCTTGTATTATATGAGAC
GTATGCACCAACGACTCTGGCTCCTGCCCGGATTTCTGGACCCTGAGATACACAACGAG
CCTCGACAATGGGAGTTTGTGGTTTGTGAGAGGTCGCTATCTGGCTCTGGAGCTGGGCC
TAATGCTGCTTCAGCTTCTCAGTTTGTGAGAGCAGAAATGCTTTCTAGTGGGTATTAAAT
AAGGCCTTGTGATGGTGGTGGTTCTATTATTCACATTGTCGATCACCTTAATCTTGAGGC
TTGGAGTGTTCGGATGTGCTTCGACCCCTTTATGAGTCATCCAAAGTCGTTGCACAAAA

AATGACCATTTCGCGGTTGCGGTATATCAGGCAATTAGCCCAAGAGTCTAATGGTGAAGT
 AGTGATGGATTAGGAAGGCAGCCTGCTGTTCTTAGAACCTTTAGCCAAAGATTAAAGCAG
 GGGCTTCAATGATGCGGTTAATGGGTTTGGTGACGACGGTGGTCTACGATGCATTGTGA
 TGGAGCGGAAGATATTATCGTTGCTATTAACTCTACAAAGCATTTGAATAATATTTCTAA
 TTCTCTTTTCGTTCTTGGAGGCGTGCTCTGTGCCAAGGCTTCAATGCTTCTCCAAAATGT
 TCCTCTGCGGTTTTGATCCGGTTCCTTAGAGAGCATCGATCTGAGTGGGCTGATTTCAA
 TGTGATGCATATTCGCGTGCTACACTTAAAGCTGGTAGCTTTGCTTATCCGGGAATGAG
 ACCAACCAAGATTCAGTGGGAGTCAGATCATAATGCCACTAGGACATACAATTGAACACGA
 AGAAATGCTAGAAGTTGTTAGACTGGAAGGTCATTCTCTTGTCTCAAGAAGATGCATTAT
 GTCACGGGATGCCATCTCCTTCAGATTTGTACCGGGATTGACGAGAATGCCGTTGGAGC
 TTGTTCTGAAGTGATATTTGCTCCGATTAATGAGATGTTCCCGGATGATGCTCCACTGT
 TCCCTCTGGATTCCGAGTCATACCCGTTGATGCTAAAACGGGAGATGTACAAGATCTGTT
 AACCGCTAATCACCGTACACTAGACTTAACTTCTAGCCTTGAAGTCGGTCCATCACCTGA
 GAATGCTTCTGGAACTCTTTTTCTAGCTCAAGCTCGAGATGTATTCTCACTATCGCGTT
 TCAATTCCTTTTGAAAACAACCTTGCAAGAAAATGTTGCTGGTATGGCTTGTGCTATGT
 GAGGAGCGTGATCTCATCAGTTCAACGTGTTGCAATGGCGATCTCACCGTCTGGGATAAG
 CCCGAGTCTGGGCTCCAAATTGTCCTCCAGGATCTCCTGAAGCTGTTACTCTTGTCTCAGTG
 GATCTCTCAAAGTTACAGTCATCACTTAGGCTCGGAGTTGCTGACGATTGATTCACTTGG
 AAGCGACGACTCGGTACTAAAACCTTCTATGGGATCACCAAGATGCCATCCTGTGTTGCTC
 ATTAAAGCCACAGCCAGTGTTTCATGTTTGCAGAACCAAGCTGGTCTAGACATGCTAGAGAC
 AACACTTGTAGCCTTACAAGATATAACACTCGAAAAGATATTCGATGAATCGGGTCGTAA
 GGCTATCTGTTCCGACTTCGCCAAGCTAATGCAACAGGGATTGCTTGCTTGCCCTTCAGG
 AATCTGTGTGTCACGATGGGAAGACATGTGAGTTATGAACAAGCTGTTGCTTGGAAGT
 GTTGTGCTGCATCTGAAGAAAACAACAATCTGCATTGTCTGCCTTCTCCTTTGTAAA
 CTGCTCTTTTGTGTGATTGCGATTGACAGAAAAGACTAATTTAAATTTACGTTAGAGAAC
 TCAAATTTTTGTTGTTTGGTGTCTCTGTTTTGTTTTTAAAATTATTTTGATCAA
 A

>G438 Amino Acid Sequence (domain in AA coordinates: 22-85)
 MEMAVANHRERSSDSMNRHLDSGKYVRYTAEQVEALERVYAECPKPSLLRRQQLIRECS
 ILANIEPKQIKVWFQNRRCRDKQRKEASRLQSVNRKLSAMNKLMEENDRLQKQVSLVC
 ENGYMKQQLTTVVNDPSCSVVTTPOHSLRDANSPAGLLSIAETLAEFLSKATGTAVDW
 VQMPGMPGPDSVGIFAISQRCNGVAARACGLVSLPEPMKIAEILKDRPSWFRDCRSLEVF
 TMTFAGNGGTIELVYMQTYAPTTLAPARDFWTLRYTTSLDNCSFVVCERSLSGSGAGPNA
 ASASQFVRAEMLSSGYLIRPCDGGGSI IHIVDHLNLEAWSVPDVLRLPLYESSKVVAQKMT
 ISALRYIRQLAQESNGEVVYGLGRQPAVLRTFSQRLSRGFNDVNGFGDDGWSTMHCDGA
 EDIIIVAINSTKHLNINISNLSFLGGVLCASMLLQNVPPAVLIRFLREHRSEWADFNV
 AYSATLKGASFAYPGMRPTRFTGSQIIMPLGHTIEHEEMLEVVRLEGHSLAQEDAFMSR
 DVHLLQICTGIDENAVGACSELI FAPINEMFPDDAPLVPSGFRVIPVDAKTG DVQDLLTA
 NHRTLDLTSSLEVGPSPENASGNSFSSSSSRCILTI AFQFPFENNLQENVAGMACQVRS
 VISSVQRVAMAI SPGISPSLGLSKLSPGSPEAVTLAQWISQSYSHHLGSELLTIDSLGSD
 DSVLKLWDHQDAILCCSLKPQPVFMFANQAGLDMLETTLVALQDITLEKIFDESGRKAI
 CSDFAKLMQQGFACLP SGICVSTMGRHVS YEQA VAWKVFAASEENNNNLHCLAFS FVNWS
 FV*

>G47 (38..472)
 CTTCTTCTTCACATCGATCATCATAACAACAAAAAATGGATTACAGAGAATCCACCGG
 TGAAAGTCAGTCAAAGTACAAAGGAATCCGTCGTCGGAAATGGGGCAAATGGGTATCAGA
 GATTAGAGTTCCGGGAACCTCGTGACCGTCTCTGGTTAGGTTCACTCTCAACAGCAGAAGG
 TGCCGCCGTAGCACACGACGTTGCTTTCTTCTGTTTACACCAACCTGATTCTTTAGAATC
 TCTCAATTTCCCTCATTGCTTAATCCTTCACTCGTTTCCAGAACTTCTCCGAGATCTAT
 CCAGCAAGCTGCTTCTAACGCCCGCATGGCCATTGACGCCGAATCGTCCACAGTACCAG
 CGTGAACTCTGGATCGGAGATACGACGACGCTATTACGAGAATGGAGCTGATCAAGTGGA
 GCCGTTGAATATTTCACTGTATGATTATCTGGGCGGCCACGATCACGTTTGATTATCTC
 GACGGTCATGATCACGTTTGATCTTCTTTTGTAGTAAGATTTTGTACCATAATCAAAACAG
 GTGTGGTGCTAAAACTTACTCAAAACAAGATTAGGTACCACAGAGAAACAATCAAATGG
 TTGTGAATATACATTATAAGGTTTTGATTAAATGTTTGTCTGATTTAGTGAAGTTTG
 GTCCATTGTATACAAATCTATTCAAGAAACCTAGCGCGAGATCATGTTTCGTGATTGAAG
 ATTGAGATTTTTAAGTATTTCGTAATATTTTTGTAAAATACAAATAAAAAAAAAAAAAA

AAAAA

>G47 Amino Acid Sequence (domain in AA coordinates: 11-80)

MDYRESTGESQSKYKGIRRRKWKWVSEIRVPGTRDLWLGSFSTAEGAAVAHDVAFFCL
HQPDSLESLSNFPHLLNPSLVSRTSPRSIQQAASNAGMAIDAGIVHSTSVNSGCGDITTTY
ENGADQVEPLNISVYDYLGGHDHV*

>G559 (89..1285)

aaagttgctagctttaatttgcgaacttactattcttatgtgtaataatcgtttgcaagg
tcgttgatttggtgataagtcagtagaaATGgataaggagaaatctccagcacctccttg
tggaggtcttctcctccatctccatcaggtcgatgctctgcattctcagaagctggtcc
cattggtcatggttcagatgctaatacgaatgagtcgatattagccgatgcttgataa
cccacctaagaagattggacatcggcgagctcattctgaaatacttactctccctgatga
tttgagctttgatagtgatcttgggtgtggttggaatgctgctgatggagcttcttctc
tgatgagactgaagaagatttgctctctatgtatcttgatggataagtttaattcttc
tgctacatcttctgcgaagttggtagccatcaggaactgcttgaaaaatgagacaat
gatgcagacaggcacaggtcaacttccaatcctcagaatacgggttaatagctctggcga
aaggccaagaatcagggcatcaacatagccaatctatggatggttcaatgaatatcaatga
gatgcttatgtcgggaatgaagatgattctgctattgatgctaagaagtctatgtctgc
tactaaacttgctgagcttgctctcattgatcctaaacgtgctaagaggatatgggcaaa
caggcagtcgcgagcagcatcaaaagaaaggaagacgagatacatatttgagcttgagag
aaaagtacagactttgcaaacagaggctacaactctctcagcccagttgacctcttaca
gagagacacaaatggcttgactggtgaaaacaatgagctgaagctgcggttacaacaat
ggagcagcaggttcacttgaggatgaactaaacgaagcactaaaggaggaaatccagca
tctgaaggtgttgactggccaagttgctccatcagcgttgaaactatgggtcggttgatc
aaaccagcagcaattctattccaacaatcagtcgaatgcaacaatcttagctgcaaaaca
gttccagcaacttcagattcattcacagaagcagcaacaacaacaacaacaacaaca
gcaacaccaacagcagcagcagcaacagcaacagtcagtttcaacagcaacagatgca
acagcttatgcagcagcggcttcaacagcaagaacaacaaatggagtaagactcaagcc
ttcacaagcccagaaagagaactGAggaatatgaatatgtcccacgtaagtgaagaggttc
tccttctgaacaattccttctcattcataaattgttggttcacatcacttgagctctc
ttggatttttaggggttttagctaacaca

>G559 Amino Acid Sequence (domain in AA coordinates: 203-264)

MDKEKSPAPPCGGLPPSPSGRCSAFSEAGPIGHGSDANRMSHDISRMLDNPPKKIGHRR
AHSEILTLDDLSFSDSLGVVGNADGASFSDTEEDLLSMYLDMDKFNSSATSSAQVGE
PSGTAWKNETMMQTGTGSTSNPQNTVNSLGERPRIHQHSQSMDSMNINEMLSGNEDD
SAIDAKKSMSATKLALALIDPKRAKRIWANRQSAARSKERKTRYIFELERKVQLQTEA
TTLAQTLTLLQRTDNLTVNNELKLRLQTMEEQQVHLQDELNEALKEEIQHLKVLTGQVA
PSALNYGSFSGSNQQQFYNNQSMQITLAAKQFQQLQIHSQKQQQQQQQQQQHQQQQQQQ
QQYQFQQQQMQQLMQRLQQQEQQNGVRLKPSQAQKEN*

>G568 (141..995)

GACCGGCTAAAGTCAAGAACCCTCTCTGAGCTCTCACCACCTTCTCTCTACTCCCTC
TCTGCGTGTAGGATACTACTAGACAATTGACAACCAAGACTAAAGCTGTGTTGTTGGTT
CACTTCTGTTCTCTTTTCCAATGTTGTCATCAGCTAAGCATCAGAGAAACCATAGACTCT
CTGCTACAAACAAGAACCAGACTCTCACCAGTTTCTTCCATTTTCCTCATCACCAT
CGTCTTCTTCTTCATCATCATCAACCTCATCATCTCTCTTTTACCTTCTCAAGACTCTC
AAGCCCAGAAGAGATCTCTTGTCAACATGGAAGAAGTTTGGAATGACATCAACCTTGCTT
CCATCCACCACCTAAACCGACACAGCCCTCATCCACAACACACAGGCAAGGTTCA
GGGGCCAAAACACCACAACCAAAACCCTAACTCAATCTTCCAAGATTTTCTCAAAGGAT
CTTTGAACCAGGAAGCAGCACCCACAAGCCAGACCACGGGTTCTGCGCCTAATGGCGATT
CCACCACGGTCACTGTTCTTTACAGCTCTCTTTTCCACCTCCTGCAACTGTTCTGAGCT
TGAATTCGGGCGCTGGCTTCGAGTTTCTCGATAACCAAGATCCTCTGTTACCTCAAAC
CTAATCTTCATACCCACCATCACCTCTCAAACGCTCATGCCTTCAACACCTCTTTCGAGG
CTCTGGTTCCATCCAGTTCTTTTGGTAAGAAAAGAGGCCAAGATTTCAATGAAGGTTCA
GGAATAGAAGACATAAGCGTATGATCAAGAACAGAGAATCTGCAGCTCGTTCCCGCGCTA
GGAAACAGGCTTATACAAACAGGTTAGAACCTTGAAGTTGCTCACTTGCAGGCAGAAAATG
CAAGACTCAAGAGACAACAAGATCAAAAAATGGCTGCAGCAATTCAGCAACCCAAAAAGA
ACACACTTCAACGGTCTTCCACAGCTCCATTTTGAAGAAATCTACAAGTCTTGTCTCTCT
TTTGGGGATTGAGATTGTCTCATGAAGAAGTGAAAAATGGCAAAAGTTGTACCCCTTTT

TTATTAGCTATAAGTATAACTAAGCCTAAAATTGTAGAACTAAGATATTGTAGGGGAAAA
 AAGAAGATGTAAACAAAAGACCCGAAAGAGAAAAGGATCTTTCAATTTCTTAAGGCAC
 AGGAACACCTGTCTCTGGGTCCTCTCTTAATGTTCTGTCGTTTTCTATGCAAACCTTTT
 TTCCTTCTGTACTAACTTATACTTGTATTCTTG

>G568 Amino Acid Sequence (domain in AA coordinates: 215-265)
 MLSSAKHQHNHRLSATNKNQTLTKVSSISSSSPSSSSSSSTSSSSPLPSQDSQAQKRSL
 VTMEEVWNDINLASIHHLNRHSPHPQHNEPRFRGQNHNNQNPNSIFQDFLKGSNLNQEP
 PTSQTTGSAPNGDSTTVTVLYSSPPPPATVLSLNSGAGFEFLDNQDPLVTSNSNLHTHH
 HLSNAHAFNTSFEALVPSSSFGKKRGQDSNEGSGNRRHKRMIKNRESAARSRAKQAYTN
 ELELEVAHLQAENARLKRQDQKMAAAIQPKNTLQRSSTAPF*

>G580 (43..747)

CCAAAAACAAAGCATTCTATGCTATTCTGTTCTGTTCTCCAATGTTGTCATCAGCAAAG
 CATAATAAGATCAACAACCATAGTGCCTTTTCAATTTCTCTTCATCATCATCATTATCA
 ACATCATCTCCCTAGGCCATAACAAATCTCAAGTCACCATGGAAGAAGTATGGAAAGAA
 ATCAACCTTGGTTCACTTCACTACCATCGGCAACTAAACATTGGTCATGAACCAATGTTA
 AAGAACCAAAACCTTAATAACTCCATCTTTCAAGATTTCCTCAACATGCCTCTGAATCAA
 CCACCACCACCACCACCACCTTCTCTTCCACCATTGTCACTGCTCTCTATGGCTCT
 CTGCTCTTCCGCTCTGCTGCTGCTCAGCTTAACTCCGGTGTGGATTGAGTTTT
 CTTGATACCACAGAAAATCTTCTTGCTTCTAACCCTCGCTCCTTTGAGGAATCTGCAAAG
 TTTGGTTGCTTGGTAAGAAAAGAGGCCAAGATTCTGATGATACTAGAGGAGACAGAAGG
 TATAAGCGTATGATCAAGAACAGAGAATCTGCTGCTCGTTCAAGGGCTAGGAAGCAGGCA
 TATACAAACGAACCTTGAGCTTGAAATTGCTCACTTGCGAGACAGAGAATGCAAGACTCAAG
 ATACAACAAGAGCAGCTGAAATAGCCGAAGCAACTCAAAACCAAGTAAAGAAAACACTA
 CAACGGTCTTCCACAGCTCCATTTTGAGAAAAATCTACTATTTCTTTTTGGGGGAGTTTC
 AAGTGTCTTCTATGAAGATGAGAAAAACAGAAAAAGTTGTACATTTTAGCTAAGTTAA
 TTTGTGGTGGTAAGTAATGTAAAGAAAAGTGTGTGTAGAAGAAAAGTGTCTAGAAAAAG
 AAAGCAACTAACTTTCTTCTTCTCTGCTTCTCTATCAACTCTTTTGACTTTTGTACT
 TTTTTCTTCTCTACTTAACCTCTATTATTGTAATGCCAAGTCAAGTCCTTATCTAGCTA
 GTACATGAGTTTCTGTTTTCACTGGTTAAGCCAT

>G580 Amino Acid Sequence (domain in AA coordinates:162-218)
 MLSSAKHNKINNHSFAFSISSSSSLSTSSSLGHNKSQVTMEEVWKEINLGLSLHYHRQLNI
 GHEPMLKNQNPNSIFQDFLNMPLNQPPPPPPSSSTIVTALYGSPLPLPPATVLSLNS
 GVGFEFLDTTENLLASNPRSFEEAKFGCLGKKRGQSDDDRGRYKRMKNRESAARS
 RARKQAYTNELELEIAHLQTENARLKIQQEQLKIAEATQNVKKTLQRSSTAPF*

>G615 (197..1252)

TTTTTCTTTTCTTTCTTTTTTGTGCTGGTGTGAGAAATGTACGCTTACTATCTCTCTCT
 CTCTCTGCCAGATTCTCTCTTTTGTATGATGTGAAAGTTGTGCTTTTGTTCCTTAAGAAA
 AAGGCATATTTTTAATACTTGATTCTTGGTTCTTGATTCTTGATTCTTGGTTTTTTTAG
 CTTCTTAAGTTCGGTGATGTCGCTCTCCACCAATGACTACAACGATGGTAATAACAATGG
 AGTGTACCTCTCTCTCTTTACCTTCTTCACTCTCTGGCCATCAAGACATCATTCATAA
 TCCCTACAACCATCAGTTAAAAGCATCTCCGGGCCATATGGTATCAGCAGTTCCTGAATC
 TCTGATCGATTACATGGCGTTTAAAGTCAAATAATGTTGTGAATCAACAAGGCTTTGAGTT
 TCCTGAGGTGTCAAAGGAAATCAAGAAGGTGGTGAAGAAGGACCGACATAGCAAGATTCA
 AACGGCACAAGGGATTAGAGACAGGAGGGTTAGGCTTTTTATTGGGATTGCTCGCCAATT
 CTTTGATCTTCAGGATATGTTGGGGTTTGATAAAGCTAGTAAAACGTTAGACTGGCTGCT
 CAAGAAGTCAAGAAAAGCCATCAAAGAGGTCGTACAAGCAAAAACCTCAACAATGATGA
 TGAAGATTTGGAAACATTGGAGGCGATGTAGAACAAGAAGAGGAGAAGGAGGAGGATGA
 CAATGGCGATAAGAGCTTCGTGTATGGTTTGAGCCCCGGGTACGGTGAAGAAGAAGTGGT
 ATGTGAGGCCACGAGCAGGGATAAGAAAGAAGAAGAGTGAGTTGAGAAACATCTCATC
 AAAGGGGCTAGGAGCCAAAGCTAGAGGAAAAGCAAGGAGCGAACAAGAGATGATGGC
 CTATGATAATCCAGAGACTGCCTCTGATATTACACAATCTGAAATCATGGACCCATTCAA
 GAGGTCTATAGTCTTCAATGAAGGAGAAGATATGACACACCTTTTCTACAAGGAACCAAT
 CGAGGAGTTTGATAATCAAGAATCTATCTTAACCAATATGACTCTACCAACGAAGATGGG
 TCAAAGTTACAATCAAAATATGGGATACTTATGTTGGTAGATCAGAGTTCTAGCAGCAA
 CTATAATACATTTCTGCCTCAAAATTTGGATTATAGTTATGATCAAAACCTTTTCATGA
 CCAAACCTTATATGTAGTCACCGACAAAATTTCCCAAGGTTTCTTATAAATCTCGAC
 AGTTTTGAAGGACTATGCATGATCAAGTTTAAACATGTAAGCCAATATAGTCCCTTATTC

CTCTGAATGTATACAAAATCTATAGTTATGTATATCTGTTCCCTTTTAAACGTATCTTTAT
 TGATCTTCTGTGCCTTGATCAAAATTGTCAATTTAAGATTAGTTTGTGTAATATTTTAG
 CTACAACCTTTTAAGTGGTATTATTGTAACCTTTTGAACATATATTTTGAAGATGAATAA
 GAACATGTTTATATAAAAA

>G615 Amino Acid Sequence (domain in AA coordinates:88-147)
 MSSSTNDYNDGNNGVYPLSLYLSSLGHQDIIHNPYNHQLKASPGHMVSAVPESLIDYM
 AFKSNNVVNQGFPEVSKEIKKVVKDRHSKIQTAGQIRDRVRVLFIGIARQFFDLQD
 MLGFDKASKTLDWLLKKSRKAIKEVVQAKNLNNDDEDFGNIGGDVEQEEKEEDDNGDKS
 FVYGLSPGYGEEVVCEATKAGIRKKKSELRNISSKGLGAKARGKAKERTKEMMAYDNPE
 TASDITQSEIMDPFKRSIVFNEGDMTHLFYKEPIEEFDNQESILTNMTLPTKMGQSYNQ
 NNGILMLVDQSSSNYNTFLPQNLDSYDQNPFDQTLVYVTDKNFPKGFL*

>G732 (73..588)

AAAAAACCAACATAAAACATAAACTCTGTCCTTTTTTGTCTTCTTGTAACCTTTTCT
 TGTTAAAAATCAATGGCGTCATCTAGCAGCACATACCGGAGCTCAAGCTCTTCCGACGGT
 GGTAAATAAATACCCGTCGACTCCGTCGTCACCGTCGACGAACGAAAACGTAAAGAATG
 TTATCGAACAGAGAATCTGCACGTAGGTCAAGGATGCGTAAACAGAAACACGTTGATGAT
 CTAACGGCTCAGATCAATCAGCTATCAAACGACAACCGTCAGATCTTGAACAGCCTCACC
 GTAACATCTCAGCTTTACATGAAGATCCAAGCCGAGAATCTGTTCTCACCCTCAGATG
 GAGGAGCTTAGCACCAGACTCCAATCTCTCAACGAGATCGTTGATCTTGTTCATCCAAC
 GGTGCAGGATTTGGTGTGACACGATCGACGGCTGTGGTTTGTATGATCGTACGGTTGGG
 ATCGACGGATATTACGATGATATGAATATGATGAGTAATGTTAATCATTGGGGTGGTTCG
 GTTTACACTAACCAACCCATTATGGCTAATGATATCAATATGTATTGATTAAATAAAATTA
 ATTAATAATAATTAGATGCCCCCTTTTTGTCTTTTTATTTTAAATTTAGCCCCATTTGGT
 GTTTTTGGGTGGTGTGATGATGTAATTATAGTACATGCATCTTTGATTGGTTGGAAGGA
 TAAATATAAACTTTATATATATATTGGGGCATATATATAGAGTTGACTTTGCATGTAT
 TGGTGTGTGTTTGTATATAATTATATGATTATATATGTTTATGTTAAAAA

>G732 Amino Acid Sequence (domain in AA coordinates: 31-91)
 MASSSTYRSSSSSDGGNNPNPSDSVTVDERKRRMLSNRESARRSRMRKQKHVDDLTAQ
 INQLSNDNRQILNSLTVTISQLYMKIQAENSVLTAQMEELSTRLOSLNEIVDLVQSNAGAF
 GVDQIDGCGFDDRTVGIDGYDDMNMSNVNHWGGSVYTNQPIMANDINMY*

>G988 (1..1338)

ATGCTTACTTCTTCAAATCCTCTAGCTCCTCCTCCGAAGATGCCACCGCTACCACCACC
 GAGAATCCTCCTCTTTGTGCATCGCTCCTCCTCGCCGCAACCTCCGCCTCACATCAC
 CTCCGTCGCTTCTTTTACCCTGTCGAATTTTCGTCTCCAGTCAAACCTCACCAGCCGCT
 CAAAACCTTACTCTCAATCCTCTCCCTTAACTCTTCTCCTCAGGCGACTCCACCGAGCGA
 CTTGTACACCTCTTCACTAAGCCTTGTCCGTACGAATCAACCGTCAGCAACAAGATCAG
 ACGGCTGAAACGGTTGCCAGTGGACGACGAACGAATGACGATGAGTAACTCCACGGTG
 TTCACGAGCAGTGTATGCAAGAAGACAGTTCTTGTTCGAACCAAGAACAACAACTCTGAC
 TTCGAGTCTTGTACTATCTTTGGCTAAACCAACTAACGCCGTTTATTTCGGTTCGGTCAT
 TTAACGGCGAACCAAGCTATCTCGACGCGACGAGACAAACGATAACGGAGCTCTACAT
 ATACTTGATTTAGATATATCACAAGGACTTCAATGGCCTCCATTGATGCAAGCCCTAGCA
 GAGAGGTCATCAAACCTAGCAGTCCACCTCCATCTCTCCGCATAACCGGATGCGGTCGA
 GATGTAACCGGATTAAACCGAACTGGAGACCGGTTAACCCGGTTCGCTGACTCTTTAGGT
 CTCCAATCCAGTTTACACGCTAGTGTATCGTAGAAGAAGATCTCGCCGGACTTTTGCTA
 CAGATCCGATTGTTAGCTCTCTCAGCCGTACAAGGAGAGACCATTGCCGTCAATTGTGTT
 CACTTCTCCACAAAATATTTAACGACGATGGAGATATGATCGGTCACTTCTTGTACGCG
 ATCAAGAGCTTAACTCTAGAATCGTTACAATGGCAGAGAGAGAAGCTAATCATGGAGAT
 CACTCGTCTTGAATAGATTCTCTGAGGCAGTGGATCATTACATGGCGATCTTTGATTGCG
 TTGGAAGCGACGTTGCCGCCAAATAGCCGAGAGAGACTAACCCCTAGAGCAACGGTGGTTC
 GGTAAGGAGATTTTGGATGTTGTGGCGGCGGAAGAGACGGAGAGAAAGCAAAGACATCGG
 AGGTTTGAGATTTGGGAAGAGATGATGAAGAGGTTTGGTTTCGTTAAGCTTCTATTGGA
 AGCTTTGCTTTGCTCAAGCTAAGCTTCTTCTTAGACTTCATTATCCTTCAGAAGGTTAT
 AATCTTCAGTTCCCTTAAACAAATCTTTGTTTCTTGGCTGGCAAAATCGTCCCCTCTTCTCC
 GTTTCGTCGTGGAATGA

>G988 Amino Acid Sequence (domain in AA coordinates:178-195)
 MLTSFKSSSSSEDATATTENPPPLCIASSSAATSASHHLRRLLFTHANFVSQSNFTAA
 QNLLSILSLNSSPHGDSRTERLVHLFTKALSVRINRQQDQTAETVATWTTNEMTMSNSTV

FTSSVCKEQFLFRITKNNNSDFESCYYLWLNQLTPFIRFGHLTANQAILDATETNDNGALH
 ILDLDISQGLQWPPLMQALAERSSNPSSPPSLRITGCGRDVTGLNRTGDRLTRFADSLG
 LQFQFHTLVIVEEDLAGLLQLIRLLALS AVQGETIAVNCVHFLHKIFNDDGDMIGHFLSA
 IKSLSNRIVTMAEREANHGDHSFLNRFSEAVDHYMAIFDSLEATLPPNSRERLTLEQRWF
 GKEILDVVAEEETERKQRHRRFÈIWEEMMKRFGFVNPIGSFALSQAKLLRLHYPSEGY
 NLQFLNNSLFLGWQNRPLFSVSSWK*

>G1519 (1..1146)

ATGAGGCTTAATGGGGATTCGGGTCCGGGTCAGGATGAACCCGGTTCGAGCGGGTTTCAC
 GGCGGAATCAGACGATTCCCGTTAGCAGCTCAGCCGAGATTATGAGAGCTGCTGAGAAA
 GACGATCAATACGCTTCTTTCATCCACGAAGCTTGCCGCGATGCCCTCCGACACCTTTTC
 GGTACAAGAATCGCTCTTGCTTACCAGAAGGAGATGAAGCTACTTGGACAGATGCTTTAC
 TATGTTCTTACGACAGGTTTACGGGCAACAACTTTAGGAGAGGAATATTGTGACATTATA
 CAGGTTGCAGGGCCTTATGGACTCTCTCTACACCAGCTAGACGTGCTTTGTTTCATATTG
 TACCAGACCGCAGTTCCATATATCGCAGAGAGAATTAGCACTCGAGCTGCTACGCAAGCA
 GTCACCTTTGATGAGTCTGATGAGTTTTTTGGTGATAGTCATATCCACTCACCAGAATG
 ATAGATCTTCCATCTTCATCTCAAGTTGAACTTCAACTTCTGTAGTATCTAGGTTAAAC
 GATAGACTTATGAGATCGTGGCACCAGCTATTACGCGATGGCCTGTGGTTCTTCTCTGTT
 GCGCGCAAGTCTTACAACCTGGTTTTGCGTGCCAATCTGATGCTCTTCTACTTTGAAGGT
 TTTTATTATCATATATCGAAACGTGCATCCGGGGTTCGTTATGTTTTTCATAGGAAAGCAA
 CTGAATCAGAGACCTAGATACCAAACTTCTGGGGTTTTCTTCTAATCCAATTGTGCATC
 CTTGCTGCTGAGGGCTTGCGTCCGAGTAATTTGTCATCTATCACTAGCTCCATTACGAG
 GCTTCTATAGGATCTTATCAAACCTTACGAGGGAGAGGTTTACCTGTTTTAAATGAAGAG
 GGAATTTGATAACTTCGGAAGCTGAAAAGGGAACTGGTCTACCTCCGATTCAACTTCA
 ACGGAGGCAGTAGGGAAATGCACTCTCTGCTTAAGCACCCGTCAGCACCCAACGGCCACT
 CCTTGTGGTCATGTGTTTTGTTGGAGCTGCATTATGGAATGGTGAACGAGAAGCAAGAA
 TGCCCTCTTTGTCGAACGCCCAATACCCATTCAAGTTTGGTTTTGTTGTATCATTCTGAT
 TTTTAG

>G1519 Amino Acid Sequence (domain in AA coordinates: 327-364)

MRLNGDSGPGQDEPGSSGFHGGIRRFPLAAQPEIMRAAEKDDQYASFIHEACRDAFRHLF
 GTRIALAYQKEMKLLGQMLYYVLTGSGQQTLGEEYCDIIQVAGPYGLSPTPARRALFIL
 YQTAVPYIAERISTRATQAVTFDESDEFFGDSHIHSPRMDLPSSSQVETSTSVVSRNLN
 DRLMRSWHRAIQRPVVLPAVAREVLQLVLRANLMLFYFEGFYHISKRASGVRYVFIGKQ
 LNQRPRYQILGVFLLIQLCILAAEGLRRSNLSSITSSIQQASIGSYQTSGGRGLPVLNEE
 GNLITSEAEKGNWSTSDSTSEAVGKCTLCLSTRQHPTATPCGHVFCWSCIMEWCNEKQE
 CPLCRTPNTHSSLVCLYHSDF*

>G374 (1..1359)

ATGGACAACAAAATGATCAGGATATTGATGTTAGATCAGTGGTTGAAGCTGTTTCCGCC
 GATCTTTCCCTTTGGTGCTCCCTCTATGTGGTTGAGAGCATGTGCATGCGCTGCCAAGAA
 AATGGAACAACAGATTCTTATGACCTTAATTCCTCAGAAAGGTTCTTAATATCT
 GCATTTGAATGTCCGCATTGCGGGGAAAGGAATAATGAAGTTCAGTTCCGAGGCGAGATT
 CAACCCCGTGATGCTGTACAATCTAGAGGTTCTAGCTGGTGATGTGAAGATATTTGAC
 CGGCAAGTTGTGAAATCTGAATCAGCCACTATTAAGATTCTGAACTGGATTTTGAGATT
 CCACCAGAGGCCCAACGTGGAAGTTTGTCTACTGTGGAAGGGATATTAGCACGGGCTGCT
 GATGAACTGAGTGCCCTTCAAGAAGAACGCAAGAAAGTTGATCCTAAACTGCTGAAGCA
 ATAGACCAATTCTTGTCCAAACCTGAGAGCTTGTGCTAAAGCAGAGACATCCTTCACCTTC
 ATTTTGGATGATCCTGTGGAACAGTTTCATTGAGAACCACATGCTCCATCACCAGAT
 CCTCTCTAACCATCAAAATCTATGAGCGAACACCAGAGCAACAAGCAACACTTGGATAT
 GTTGCTAACCATCTCAGGCTGGACAATCAGAAGGAAGCCTTGGCGCACTGTGATGACT
 TTCCCTTCAACTTGGCGAGCATGTACGGAGCCGTGTGAGACACGGATGTTCAAAATAGAA
 ATCCCGTACTTTTTCAGGAAGTTATTGTGATGGCATCTACATGTGACAGTTGTGGCTATCGT
 AATTCTGAGTTGAAGCCTGTTGGTGCAATTCCTGAAAAGGGAAAGAGATTACTCTCTCT
 GTGAGGAACATTACAGACCTTAGCCGAGATGTTATCAAGTCGGACACTGCAGGAGTGATA
 ATCCCAGAACTTGATCTGGAGCTAGCTGGTGGTACACTTGGTGGAAATGGTAACAACAGTT
 GAAGGGTTGGTTACACAGATCAGAGAAAGCCTAGCGAGAGTTCACGGATTCACTTTTGGT
 GATAGTATGGAAGAGTAAGTTGAACAAATGGAGAGAATTTGGAGCCAGGCTCACTAAG
 CTCCTAAGCTTTGAACAGCCGTGGCAATTGATTCTTGATGATGAATTAGCAAATTCCTTT
 ATTGCACCAGTAACAGATGATATCAAAGATGACCATCAGCTCACATTTGAAGAGTACGAG

BNSDOCID: <WO__03013227A2 I_>

MAGFDENVAVMGEWVPRSPSPGTLFSSAIGEEKSSKRVLERELSLNHGQVIGLEEDTSSN
 HNKDSSQSNVFRGGLSERIAARAGFNAPRLNTENIRNTDFSIDSNLRSPCLTISSPGLS
 PATLLESFVFLSNPLAQSPSTTGKFPFLPGVNGNALSSEKAKDEFDDIGASFSPHVS
 SSSSFQGTTEMMSVDYGNYNRRSSSHQSAEEVKPGSENISSNLYGIETDNQNGQNKTS
 DVTNTSLETVDHQEEEEQRRGDSMAGGAPAEADGYNWRKYGQKLVKGSEYPRSYKCTN
 PNCQVKKKVERSREGHITEIIYKGAHNHLKPPNRRSGMQVDGTEQVEQQQQQORDSAATW
 VSCNNTQQQGSNNENVEEGSTRFEYGNQSGSIQAQTGGQYESGDPVVVVVDASSTFSNDE
 DEDDRGTHGSVSLGYDGGGGGGGGEGDESESKRRKLEAFAAEMSGSTRAIREPRVVVQTT
 SDVDILDDGYRWRKYGQKVVVKGNPNPRSYKCTAPGCTVRKHVERASHDLKSVITTYEGK
 HNHDVPAARNSSHGCGGDSGNGNSGGSAVSHHYHNGHHSEPPRGRFDRQVTTNNQSPFS
 RPFSPQPHLGPPSGFSGFLGQTGLVNLMPGLAYGQGKMPGLPHPYMTQPVGMSEAMMQR
 GMEPKVEPVSDSGQSVYNQIMSRLPQI*

>G1000 (1..954)

ATGGGAAGACCTCCTTGTGTGACAAGTCCAATGTCAAGAAAGGTCTCTGGACCGAGGAA
 GAAGACGCTAAGATCCTTGCTTATGTTGCTATCCATGGTGTAGGAACTGGAGCTTGATC
 CCCAAAAAGCAGGTCTGAATCGATGTGGAAAGAGCTGTAGACTAAGATGGACTAATTAC
 TTAAGACCTGACCTTAAACATGACAGCTTCTCTACCAAGAAGAAGAGCTTATCATTGAG
 TGTCATAGAGCCATTGCGCAGCAGGTGGTCTTCCATTGCACGAAAGCTTCCAGGAAGAACG
 GATAATGATGTGAAGAATCACTGGAACACAAAGCTGAAGAAGAAGCTGATGAAAATGGGG
 ATAGACCCGGTGACTCATAAACCGGTTTCTCAACTCCTTGCAGAATTCAGAAACATTAGC
 GGCCATGGAAATGCATCCTTCAAAACAGAACCATCTAACAACTCTATACTCACACAATCC
 AACTCAGCTTGGGAAATGATGAGAAACACAACAACAAACCATGAGAGTTATTACACCAAC
 TCTCCAATGATGTTTACAAATTCCTCTGAGTACCAAACTACTCCATTTTCATTTCTATAGC
 CATCCAAATCATCTGCTCAATGGAACCACATCTTCATGCTCTTCTCATCATCTTCTACT
 AGTATCACTCAGCCAAACCAAGTACCTCAAACACCGGTTACTAACTTCTACTGGAGCGAT
 TTCCTTCTCTCGGACCCGGTTCCTCAAGTAGTGGGATCCTCAGCTACTAGCGACCTCACT
 TTTACGCGAAGCAACATCATTTCAACATCGAAGCCGAATACATCTCTCAAAACATCGAT
 TCAAAGGCTCGGGAACATGTCATTCCGCGAGTTCCTTCGTTGACGAAATACTAGATAAA
 GACCAAGAGATGTTGTACAGTTTCTCTCAACTCTTGAATGATTTGATTATTAG

>G1000 Amino Acid Sequence (domain in AA coordinates: 14-117)

MGRPPCCDKSNVKKGLWTEEDAKILAYVAIHGVGNWSLIPKAGLNRCGKSCRLRWTNY
 LRPDLKHSFSTQEEELIIECHRAIGSRWSSIARKLPGRTDNDVKNHWN TKLKKKLMKMG
 IDPVTHKPVSQLLAEFRNISGHGNASFKTEPSNNSILTQSN SAWEMMRNTTNNHESYYTN
 SPMFTNSSEYQTTTFHFYSHPNHLLNGTTSSCSSSSSTSITQPNQVPQTPVTNFYWS
 FLLSDPVPQVVGSSATS DLTFTQNEHHFNIEAEYISQNI DSKASGTCHSASSFVDEILD
 K DQEMLSQFPQLLNDFDY*

>G1067 (436..1371)

TCTCAAGCTTCTCTCTCTCTTTTCTCCCATAGCACATCAGAATCGCTAAATACGACTCCT
 ATGCAAGAAGAAGCTACTTCTTCTCTTGCCCTAATTAATCTACCTAAGGTTTCC
 TCTTACCTTTCATGAGAGAGATCATTTAACATAAGTCACCTTTTTTATATCTTTTGCTTC
 GTCTTTAATTTAGTCTGTCTTGGTCTGTTTCTATATTTGTGCGCTTGCGTAACCGAT
 CACACCTTAATGCTTTAGCTATTGTTTCTCAAATCATGAGTTTGTACTTCTCGATCTG
 AGTTTTCTTTTCTCTCTTTACGCTCTTCTTACCTAGCTACCAATATATGAACGAGCAG
 GATCAAGAATCGAGAAATTGATTTGAGCTGGCGAATAAGCAGTGGTGGGATAGGGAATTA
 GTAGATGCGGCGGCGATGGAAGGCGGTACGAGCAAGGCGGTGGAGCTTCTAGATACTTC
 CATAACCTCTTTAGACCGGAGATTCAACCAACAGCTTCAACCGCAGGGCGGGATCAAT
 CTTATCGCAGCATCATCAGCACCAGCAACATCAACAACAACAACCGTCGGAT
 GATTCAAGAGAATCTGACCATTCAAACAAGATCATCATCAACAGGGTCGACCCGATTCA
 GACCCGAATACATCAAGCTCAGCACCGGGAACCGTCCACGTGGACGTCCACCAGGATCT
 AAGAACAAGCCAAGCCACCGATCATAGTAACTCGTGATAGCCCCAACGCGCTTAGATCT
 CACGTTCTTGAAGTATCTCCTGGAGCTGACATAGTTGAGAGTGTTCACGTACGCTAGG
 AGGAGAGGGAGAGCGTCTCCGTTTGTAGGAGGAAACGGCACCCTATCTAACGTCACTCTC
 CGTCAGCCAGTCACTCTGGAATGGCGGTGGTGTGTCGGAGGAGGAGGAGTTGTGACT
 TTACATGGAAGGTTTGTGATTTCTTCGCTAACGGGACTGTTTGGCCACCTCTGCACCG
 CCTGGTGCCGTTGTTGTCTATATTTTAGCCGGAGGGCAAGGTGAGTGGTGGGAGGA
 AGCGTTGTGGCTCCCTTATTGTCATCAGCTCCGGTTATATACTAATGGCGGCTTCGTTCTCA
 AATGCGGTTTTTCGAGAGACTACCGATTGAGGAGGAGGAAGAAGAAGGTGGTGGTGGCGGA

GGAGGAGGAGGAGGAGGGCCACCGCAGATGCAACAAGCTCCATCAGCATCTCCGCCGTCT
GGAGTGACCGGTCAGGGACAGTTAGGAGGTAATGTGGGTGGTTATGGGTTTTCTGGTGAT
CCTCATTTGCTTGGATGGGGAGCTGGGAACACCTTCAAGACCACCTTTTAAATTGAATTTT
AATGTCCGGAATTTATGTGTTTTATCATCTTGAGGAGTCGTCTTTCCTTTGGGATATT
TGGTGTTTAATGTTTAGTTGATATGCATATTTT

>G1067 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEGGYEQGGGASRYFHNLFPEIHHQQLQPQGGINLIDQH HHQHQQHQQQQPSDDSRES
DHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVLEV
SPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTTPGNGGGVSGGGGVVTLHGRF
EILSLTGTVLPPPAPPAGGLSIFLAGGQGVVGGSVVAPLIASAPVILMAASFNAVFE
RLPIEEEEEGGGGGGGGGPPQMQQAPSASPPSGVTGQQLGGNVGGYGFSGDPHLLG
WGAGTPSRPPF*

>G1075 (19..876)
TTTGTGTTTGGTGTGGCATGGCTGGTCTCGATCTAGGCACAACCTTCTCGCTACGTCCAC
AACGTCTGATGGTGGCGGCGGCGGACAGTTCACCACCGACAACCACCACGAAGATGACGGT
GGCGCTGGAGGAAACCACCATCATCACCATCATAATCATAATCACCATCAAGGTTTAGAT
TTAATAGCTTCTAATGATAACTCTGGACTAGGCGGCGGTGGAGGAGGAGGGAGCGGTGAC
CTCGTCATGCGTCGGCCACGTGGCCGTCCAGCTGGATCGAAGAACAACCGAAGCCGCCG
GTGATTGTACGCGCGAGAGCGCAAACTCTTAGGGCTCACATTCTTGAAGTTGGAAGT
GGCTCGGACGTTTTTCGAATGTATCTCCACTTACGCTCGTCGGAGACAGCGCGGATTTGC
GTTTTATCCGGGACGGGAACCGTCACTAACGTGAGCATCCGTGAGCCTACGGCGCGCGGA
GCTGTTGTGACTCTGCGGGGTACTTTTGAGATTCTTCCCTCTCCGGATCTTTCTTCCG
CCACCTGCTCCTCCAGGGGCGACTAGCTTGACGATATTCTCGCTGGAGCTCAAGGACAG
GTCGTCGGAGGTAACGTAGTTGGTGAGTTAATGGCGGCGGGCCGGTAATGGTCATGGCA
GCGTCTTTTACAACTGCGCTTACGAAAGGTTGCCTTTGGACGAGCATGAGGAGCACTTG
CAAAGTGGCGGCGCGGAGGTGGAGGGAATATGTACTCGGAAGCCACTGGCGGTGGCGGA
GGGTTGCCTTTCTTAAATTTGCCGATGAGTATGCCTCAGATTGGAGTTGAAAGTTGGCAG
GGGAATCAGCGCGCGCGCGGTAGGGCTCCGTTTTAGCAATTTAAGAACTTTAATTGTTT
TTTCCACTTTTTTTGTTTTCTCCGAATTTTATGAAATTATGATTTAAGAAAAAACCAGAT
ATTGTTTCATGTATTGACCCTCTTACTGCATGGTTTTCTTCTATTGGGTAAATTGGCTAGCT
CATAAGAATTGTTTAAATTTGGTTATTGTTCATCAAATTTGCCACATATAAAGCTTCTAGC
AAAT

>G1075 Amino Acid Sequence (domain in AA coordinates: 78-85)
MAGLDLGTTSRYVHNVDGGGGGQFTTDNHHEDDGGAGGNHHHHHHNHNHHQGLDLIASND
NSLGGGGGGGSGDLVMRRPRGRPAGSKNPKPPVIVTRESANTLRAHILEVSGCDVFE
CISTYARRRQRGICVLSGTGTVTVNSIRQPTAAGAVVTLRGTFEILSLSGSFLPPPAPP
ATSLTIFLAGAQGVVGGNVVVGELMAAGPVMVMAASFNTVAYERLPLDEHEEHLQSGGGG
GGGNMYSEATGGGGGLPFFNLPMSPQIGVESWQGNHAGAGRAPP*

>G1266 (62..718)
CAATCCACTAACGATCCCTAACCGAAAACAGAGTAGTCAAGAAACAGAGTATTTTTCTA
CATGGATCCATTTTAAATTCAGTCCCCATTTCTCCGGCTTCTACCGGAATATTCTATCGG
ATCTTCTCCAGATTCTTCTCATCCTCTTCTTCTAACAATTAATCTCTTCCCTTCAACGA
GAACGACTCAGAGGAAATGTTTCTCTACGGTCTAATCGAGCAGTCCACGCAACAAACCTA
TATTGACTCGGATAGTCAAGACCTTCCGATCAAATCCGTAAGCTCAAGAAAGTCAGAGAA
GTCTTACAGAGGCGTAAGACGACGCGCATGGGGGAAATTTCGCGCGCGAGATAAGAGATT
GACTAGAAACGGTATTAGGGTTTGGCTCGGGACGTTTCAAAGCGCGGAAGAGGCGGCTTT
AGCCTACGATCAAGCTGCTTTCTCGATGAGAGGGTCTCGGCGATTCTCAATTTTTCGGC
GGAGAGAGTTCAAGAGTCGCTTTTCGAGATTAAATATACCTACGAGGATGGTTGTTCTCC
GGTTGTGGCGTTGAAGAGGAAACACTCGATGAGACGGAGAATGACCAATAAGAAGACGAA
AGATAGTGACTTTGATCACCGCTCCGTGAAGTTAGATAATGTAGTTGTCTTTGAGGATTT
GGGAGAACAGTACCTTGAGGAGCTTTTGGGGTCTTCTGAAATAGTGGGACTTGGTGAAA
GATTAGGATTTGTATTAGGGACCTTAAGTTTGAAGTGGTTGATTAATTTTAAACCTAATA
TGTTTTTTGTTGCTTAAATATTGATTCTATTGAGAAACATCGAAAACAGTTTGTATGT
ACTTTTGTGATACTTGGCG

>G1266 Amino Acid Sequence (domain in AA coordinates: 79-147)
MDPFLIQSPFSGFSPEYSIGSSPDSFSSSSNNYSLPFNENDSEEMFLYGLIEQSTQQTY
IDSDSQDLPIKSVSSRKSEKSYRGVRRRPWGKFAAEIRDSTRNGIRVWLGTFFESABEEAL

AYDQAAFSMRGSSAILNFS AERVQESLSEIKYTYEDGCS PVVALKRKHS MRRRMTNKKTK
DSDFDHRSVKLDNVVVFEDLGEQYLEELLGSSENSGTW*

>G1311 (41..757)

AAGTATAATAACACAAAGAAACAGAGTAAAGAAAGAAAAATGGATTTTAAGAAGGAAGA
AACACTTCGTAGAGGGCCATGGCTCGAAGAAGAAGACGAACGGCTAGTGAAGGTCATTAG
TCTTTTGGGAGAACGTCGTGGGATTCTTTAGCAATAGTTTCCGGTTTGAAGAGGAGTGG
TAAGAGTTGCAGGCTAAGGTGGATGAACATCTGAATCCGACTCTGAAGCGTGGACCGAT
GAGTCAAGAAGAAGAGAGAATCATCTTTCAGCTCCATGCTCTATGGGGTAACAAGTGGTC
GAAGATTGCGAGAAGATTACCCGGTAGGACTGATAACGAGATAAAGAACTATTGGAGAAC
TCATTATAGAAAGAAACAGGAAGCTCAAACTATGGAAAGCTCTTTGAGTGGAGAGGAAA
TACAGGAGAAGAATTGTTGCACAAGTATAAGGAAACAGAGATCACTAGGACAAAGACGAC
GTCTCAAGAACATGGTTTTGTGTAAGTTGTGAGCATGGAAGTGGTAAAGAAGCCAACGG
TGGTGTGGTGGGAAGAGAAAGCTTCGGTGTATGAAATCACCGTATGAAAATCGGATTTC
GGATTGGATATCAGAGATTCTACTGACCAGAGTGAAGCAAATCTTTCAGAAGATCACAG
CAGCAATAGCTGCAGTGAGAACAATATTACATTGGTACTTGGTGGTTTCAAGAGACTAG
GGACTTTGAGGAGTTTTTCATGTTCTCTATGGTCATAATTCTAAAGTTGGTTTATTACTT
TTTAAAAA

>G1311 Amino Acid Sequence (domain in AA coordinates: 11-112)

MDFKKEETLRGPWLEEDERLVKVISLLGERRWDSLAIIVSGLKRSKSCRLRWMNYLNP
TLKRGPMSEQEERIIFQLHALWGNKWSKIARRLPGRTDNEIKNYWRTHYRKKQEAQNYGK
LFEWGRNTGEELLHKYKETEITRKTTSQEHGFVEVVSMSGKEANGVGGRSFVGMKS
PYENRISDWISEISTDQSEANLSEHDHSSNSCENNINIGTWWFQETRDFFESCCLWS*

>G1321 (72..803)

GTTCTTGATTGGTTGGATCGGTATACCTTAGTTGATTACGTAATTAAATAGATCGGCGT
GAAGAAGAAAAATGATCATGTGCAGCCGAGGCCATTGGAGACCAGCTGAAGACGAGAAGC
TCAAGGATCTTGTGCAACAATACGGTCCTCACAATTGGAACGCCATTGCTCTCAAGCTTC
CTGGTCTGCTCGGTAAGAGTTGTAGATTGAGATGGTTTAATCAATTGGATCCAAGGATCA
ACCGAAACCTTTTCACGGAAGAAGAAGAAGAAAGACTTTTAGCGGCTCATCGGATCCATG
GGAACAGATGGTCCATCATCGCAAGGCTTTTCCCTGGAAGAACTGATAACGCCGTCAGA
ACCATTGGCACGTCATCATGGCTCGTCGCACACGCCAAACCTCTAAGCCTCGTCTTCTTC
CCTCGACGACTTCGTCTTCTTCTTTAATGGCGAGTGAACAAATCATGATGAGTTCTGGTG
GTTATAATCATAATTATAGTTCCGATGATCGGAAGAAAATATTTCAGCAGACTTTATAA
ATTTCCCTTACAAATTCTCTCATATCAATCATCTTCACTTCTTAAAGGAGTTTTCCTCCG
GAAAGATCGCTTTAAGTCACAAAGCAAATCAGAGTAAGAAGCCTATGGAGTTCTACAATT
TTCTACAAGTAAACACAGATTCAAACAAGAGCGAGATTATAGATCAAGATTCAAGTCAA
GCAACCGCAGTGACTCGGACACCAAACATGAAAGTCATGTTCCATTCTTCGACTTTTAT
CCGTTGGAAGCTCTGCCTCCTAGGATTAGTTTTTTTGCAGTAACTCCTAAATTTCTAGAT
TAACTATTTAGTCCGTATACGTACGAGATTATCTAGGTCGTTAGCATGTATGCTTGATGT
GTATAATCACTAACTAGTGAGCTATTACCTGCGAAAATGTAAAGAAAATACATAATGTT
GATGTATCACACATTCTCAATGTCTGTAAATTTCCATCGAGTTGTTAACTATCAAAGTT
ATCCGTTTGAAAAA

>G1321 Amino Acid Sequence (domain in AA coordinates: 4-106)

MIMCSRGHWRPAEDEKLKDLVEQYGPHNWNIAIALKLPGRSKSCRLRWFNQLDPRINRNP
FTEEEERLLAAHRIHGNRWSIIARLFPGRTDNAVKNHWHVIMARRTRQTSKPRLLPSTT
SSSSLMASEQIMSSGGYNHNYSSDDRKKIFPADFINFPYKFSHINHLHFLKEFFPGKIA
LSHKANQSKKPMEFYNFLQVNTDSNKSEIIDQDSGQSKRSDSKHESHVPFFDFLSVGN
SAS*

>G1326 (32..784)

CGACGGTACGGTGGAGATAGAGATAGCATCCATGGAGATGTCTAGAGGAAGCAACAGTTT
TGACAATAAGAAGCCTAGTTGCCAAAGAGGTCACTGGAGACCTGTTGAAGATGACAATCT
CCGGCAACTCGTTGAACAATACGGTCCCAAGAACTGGAATTTTATTGCTCAACATCTCTA
TGGAAGATCAGGGAAGCTGTAGATTAAGATGGTACAACCAACTTGATCCAAACATCAC
CAAGAAACCTTTCACCGAGGAGGAAGAAGAGAGACTGCTTAAAGCTCATCGGATCCAAGG
GAATCGTTGGGCCTCCATAGCCCGACTGTTCCCGGGAGGACCGACAACGCTGTCAAAAA
CCATTTTCATGTATCATGGCTAGACGCAACCGGGAAGCTTCTCTTCCACAGCTACTTC
TACGTTCAACCAAACTTGGCATACTGTTTGGAGCCCTAGTTCTAGTCTTACAAGGCTAAA
TAGATCCCATTTGGGCTATGGAGGTATCGAAAGGATAAGAGTTGCGGTCTCTGGCCTTA

CTCTTTTGTTCACCACTACGAATGGTCAATTTGGATCTTCATCTGTCTCTAACGTACA
 CCACGAAATTTATCTTGAGAGGAGAAAGTCGAAAGAGTTGGTGGATCCTCAGAATTACAC
 ATTTTCATGCAGCCACACCATGATCATAAGATGACTTCAAATGAAGATGGACCATCCATGGG
 AGATGATGGTGAGAAGAACGATGTTACTTTTCATTGATTTTCTTGGTGTGGATTAGCTTC
 TTAGGTTATAACATCACAAGTCAAAGCTTTTAAGGGTTTCTATCATTAGGGTTAGGCATC
 ATTTTCAGCCTTTTGCTTCCTTAACTCTCATATGGATCT

>G1326 Amino Acid Sequence (domain in AA coordinates: 18-121)
 MEMSRGSNSFDNKKPSCQRGHRPVEDDNLRLQVLEQYGPKNWNFIAQHLYGRSGKSCRLR
 WYNQLDPNITKKPFTEEEERLLKAHRIQGNRWASIALRFPGRTDNAVKNHFHVMARRK
 RENFSSSTATSTFNQTHWTVLSPSSSLTRLNRSHFGLWRYRKDKSCGLWPYSFVSPPTNGQ
 FGSSSVSNVHHEIYLERRKSKELVDPQNYTFHAATPDHKMTSNEDGPSMGDDGEKNDVTF
 IDFLGVGLAS*

>G1367 (128..1567)

TCCTTCCACAAAACCTTTTTTAAATTTTATCTGAAAAATTAAAAACAACCGAAACAAAAA
 AAAACTAAAAATCAAAAATCTCATCACCTTCCTTGCTCTGTATTTTTTCTCTCTCACTAA
 ATCCTCCATGGATCCTTCTCTCTCTGCAACCAATGATCCTCATCATCCTCCTCCTCCTCA
 GTTCACATCTTTCCCTCCTTTTACCAACACCAACCCCTTCGCCTCTCCAAACCAACCCCTT
 CTTACCCGACCCACCGCCGTCGCGCCGCCAAACAACATCCATCTCTATCAAGCAGCTCC
 TCCGACAGCCACAAACATCTCCAGTTCCTCCTCATCCATCTATTTCCACCCCTCCTTA
 CTCTGACATGATTGTCACGGCGATTGTCAGCGTTAAACGAACAGATGGGTCAAGCAAGCA
 AGCTATTTTCAGAGTACATAGAGAGAATTTACACTGGGATTCTACTGCTCATGGAGCTTT
 GTTGACACACCATCTCAAGACTTTGAAGACCAGTGGGATTCTTGTCATGGTTAAGAAATC
 TTACAAGCTTGCTTCTACTCCTCCTCCTCCTCCTACTAGTGTAGCTCCTAGTCTTGA
 ACCTCCCAGATCTGATTTTCATAGTCAACGAGAACCAACCTTTACCTGATCCGGTTTTGGC
 TTCTTCTACTCCTCAGACTATTAACCGTGGTTCGTGGTTCGACCTCCAAAAGCTAAACAGA
 TGTGTTTCAACCTCAACCTCTGACTAATGGAACCTCACCTGGGAACAGAGTGAATTACC
 TGTCTCTCGACAGAGGAGATACAGATACAGCCGCCACAGTTACCGTTACAGCCACAGCA
 GCCGGTTAAGAGACCCCGGGTCGTCTAGAAAAGATGGAACCTTCGCCGACGGTGAAGCC
 AGCTGCTTCTGTTTTCCGGTGGTGTGGAGACTGTGAAACGAAGAGGTAGACCTCCGAGTGG
 AAGAGCTGCTGGGAGGGAGAGAAAAGCCTATAGTAGTCTCAGCTCCAGCTTCAGTGTCCC
 GTATGTTGCTAATGGTGGTGTAGACGCCGAGGGAGACCAAAGAGAGTTGACGCTGGTGG
 TGCTTCTCTGTTGCTCCACCACCACCACCACCACTAACGTAGAGAGTGGAGGAGAGGA
 GGTTGCAGTCAAGAAACGAGGAAGAGGACCGCCTCCTAAGATTGGAGGTGTTATCAGGAA
 GCCTATGAAGCCGATGAGAAGCTTTGCTCGTACTGGAACCCGTTAGGAAGACCCAGAAA
 GAATGCGGTGTCTAGTGGGAGCTTCTGGACGACAAGATGGTGAATGAGAGAACTGAAGAA
 GAAGTTTGAGTTGTTTCAAGCGAGAGCTAAGGATATTGTAATTGTGTTGAAATCCGAGAT
 AGGAGGAAGTGGAAATCAAGCAGTGGTTCAAGCCATACAGGACCTGGAAGGGATAGCAGA
 GACAACAAACGAGCCAAAGCACATGGAAGAAGTGCAGCTGCCAGACGAGGAACACCTTGA
 AACCGAACCAAGCAGAGGGTCAAGGACAGACAGAAGCAGAGGCAATGCAAGAAGCTCT
 GTTCTAAAGATAAAGCCTTGACATAAAAAGCTAGCAAGTGGTGGGTTTACTTGTGTGTG
 TTACATGAAATTTTAACTTTATAAGGGTGTGTCAGGAGAAAAACAAAAGAACAATGT
 GATGAACTGATGATGATGATTGTGTCTCTAACCACCAACAAGGAGAGGTAGGGTAATGT
 CTGTAAAGTGAATTAGGATGTTACCATGTTTCATGCTTCCCATCTCTCCATCGTCCAT
 ATCTGTGTAGGCAGCTTTGTTCTTGTTCCTCGTGTGTTTTTTTAGACTGTTGTGTCTCT
 TATCTATTTGTCTCCTTAGGCTTTTTTAGGAGTTGTTGTTGATGTTTATCAAAAACGCT
 TATGTAATTTTATGACCACTTCTACTTTTTTATGATGGTTTTCTT

>G1367 Amino Acid Sequence (domain in AA coordinates: 179-201, 262-285, 298-319, 335-357)

MDPSLSATNDPHHPPPPQFTSFPPFTNTNPFASPNHPFFTGPTAVAPPNNIHLHYQAAPPQ
 QPQTSPPVPHPSISHPPYSMDICTAIAALNEPDGSSKQAIISRYIERIYTGIPAHGALLT
 HHLKTLKTSGLIVMVKKSXYKLASTPPPPPTSVAPSLEPPRSDFIVNENQPLPDPVLASS
 TPQTIKRGRGRPPKAKPDVQPPQLTNGKLTWEQSELVSRPEEIQIOPPLPLQPPQPV
 KRPPGRPRKDGTSPTVKPAASVSGGVETVKRRGRPPSGRAAGRERKPIVVSAPASVFPYV
 ANGGVRRRGRPKRVDAGGASSVAPPPPPPTNVESGGEEVAVKKRGRGRPPKIGGVIRKPM
 KPMRSFARTGKPVGRPRKNAVSVGASGRQDGDYELKKKFELFQARAKDIVIVLKSEIGG
 SGNQAVVQAIQDLEGIAETTNEPKHMEEVQLPDEEHLETEPEAEGQGQTEAEAMQEALF*

>G1386 (89..673)

AATTTTATTTCTCTCTCAAATCTTCCCACCAAAAATTAACCTTTTCGTTTCACTAAG
TCCCTTTTAAAGAAAATATCCCAATTAATGGAACGTGACGACTGCCGGAGATTTTCAGGA
CTCGCCGGCGCAGACGACGAGAGAGAAGAGTGAAATATAAACCAAGAAGAAAAGAGCCAA
AGATGATGATGATGAGAAAAGTTGTTTTCGAAGCATCCAAATTTTCGAGGTGTCAGAATGAG
ACAAATGGGGAAAATGGGTGTCCGAAATCAGAGAGCCAAAAAGAAATCAAGAATCTGGCT
CGGTACTTTCTCCACGGCGGAGATGGCGGCGGTGCTCACGACGTGGCAGCTTTAGCCAT
CAAAGGCGGTTCTGCACATCTCAACTTCCCGAGCTCGCTTATCACCTCCCTAGACCAGC
TAGTGCCGACCCCTAAAGACATCCAAGCTGCCGCCGCCGAGCTGCAGCCGCTGTGGCCAT
TGACATGGATGTAGAGACGTCTTCGCCGTGCCATCTCCACAGTTACGGAAACGTCATC
TCCGGCTATGATAGCACTCTCCGACGACGCGTTCTCCGATCTTCCTGATCTCTTGCTCAA
CGTGAACCATAACATCGATGGCTTCTGGGACTCTTTTCCCTATGAAGAACCCCTTCCTCTC
TCAAAGTTACTAGAAATCAAACTATGTCGTTTTTGTATGTATTTTGTATGTGACCA
TTTTTTGACGTGCAAAATCACCCGGATAATCCAAATTGTATGATTTATTAATGGTTGATG
ATTTCTTTGTGTGGAACAATGTGTATGATACGTAATCAAAAGTTCAAAAAAAAATAAA
AAAAA

>G1386 Amino Acid Sequence (domain in AA coordinates: TBD)
MERDDCRRFQDSPAQTERRVKYKPKKKRAKDDDEKVVSKHPNFRGVRMRQWGKWVSEI
REPKKKSRIWLGTFSTAEMAARAHDAALAIKGGSAHLNFPDELAYHLRPPASADPKDIQA
AAAAAAAVIDMDVETSSPSPPTVTETSSPAMIALSDDAFSDLPDLLLNHNIDGFW
DSFPYEEFPLSQSY*

>G1421 (292..1155)

GAAATTTTCATCCCTAAATAAGAAAAAGCATCTCCTTCTTTAGTGTCTCTTCACCAAA
CTCTTGATTCCATAAGCATATATAAAAAGCTCTCTGCTTTCTTCAACTTTCCCGGGAA
AATCTTCTTGTACAAAGCATCAATCTCTTGTTTTACCAATTTCTCTCTTTATTCTCTT
TTTGCCCTTTACTTTTCTTAACCTTTGGTCTTTTATATATAAACACACGACACAAAGAAGAA
CACACATAAGTTAAACTATTACAACAGTTTTTAAAGAGAGAGATTAAAAAATGGAGACA
GAGAAGAAAGTTTCTCTCCCAAGAAATCTTACGAATCTCTGTTACTGATCCTTACGCAACA
GATTTCGTAAGCGACGAAGAAGAAGAGTTGATTTTGTATGCATTATCTACAAAACGACGT
CGTGTTAAGAAGTACGTGAAGGAAGTGGTGTGATTCGGTGGTTTCTGATAAAGAGAAG
CCGATGAAGAAGAAGAGAAAGAGCGCGTTGTTACTGTTCCAGTGGTTGTTACGACGGCG
ACGAGGAAGTTTCGTGGAGTGAGGCAAAGACCGTGGGGAAAATGGGCGGCGGAGATTAGA
GATCCGAGTAGACGTGTAGGGTTTGGTTAGGTACTTTTGACACGGCGGAGGAAGCTGCC
ATTGTTTACGATAACGCAGTATTTCAGCTACGTGGTCTTAACGCAGAGCTTAACCTCCCT
CCTCCTCCGCTGACGGAGAATGTTGAAGAAGCTTCGACGGAGGTGAAAGGAGTTTCGGAT
TTTATCATTTGGCGGTGGAGAATGCTCTTCGTTCCGCCGTTTCTGTTCTCGAATCTCCGTTT
TCCGGCGAGTCTACTGCGGTTAAAGAGGAGTTTGTGCGGTGTATCGACGGCGGAGATTGTG
GTTAAAAAGGAGCCGTCTTTTAAAGGTTTCAAGATTTCTCGGCGCGGTTGTTCTCGGACGAC
GACGTTTTTGGTTTCTCGACGTGATGAGTGAAAGTTTCGGCGGCGGATTATTTGGAGAT
AATCTTTTTCGGATATGAGTTTGGATCCGGGTTTGGATTCCGGTCTGGGTCTGGATTCT
TCCAGCTGGCAGCTTGAGGACCATTTTCAAGATATTGGGGATTATTCGGGTCCGATCCT
GTCTTAACTGTTTAAAGAAATAACTGGCCGTTTAAACGGCGTTTGTAGTGAAGTTTGTACCG
GCGACGGCGGAGATTAAAAAAAACGGCGATTATTTTGAATGAAGATTGTTAAATA

>G1421 Amino Acid Sequence (domain in AA coordinates: 74-151)
METEKKVSLPRILRISVTDPYATDSSSDEEEVDFDALSTKRRRVKKYVKEVVLDSVSD
KEKPMKKKKRKRVTVPVVVTTATRKFRGVRQRPWKWAAEIRDPSRRVRVWLGTFTDAE
EAAIVYDNAAIQLRGPNAELNFPPTVTENVEEASTEVKGVSDFIIGGGECLRSPVSVLE
SPFSGESTAVKEEFVGVSTAEIVVKKEPSFNGSDFSAPLFSDDVDFGFSTSMSESFGGDL
FGDNLFADMSFGSGFGFGSGSFSSWHVEDHFQDIGDLFGSDPVLTV*

>G1453 (39..917)

CGTCGACGCGAAATCCTAGAAAATAACTATCAATATGATGAAGGTTGATCAAGATT
ATTCTGTAGTATACCGCTGGATTTAGGTTTCATCCGACAGATGAAGAAGTTGTGCGGAT
ATTATCTCAAGAAGAAAATCGCCTCCAGAGGATTGATCTCGACGTTATCAGAGAAATTG
ATCTTTACAAGATCGAACCATGGGATCTACAAGAGAGATGTAGGATAGGGTACGAGGAGC
AAACGGAGTGGTATTTCTTCAGCCATAGAGACAAGAAGTATCCGACTGGGACTAGGACAA
ACCGAGCCACCGTGGCCGTTTCTGGAAAGCAACGGGCCGGGACAAGGCGGTTTACCTCA
ACTCCAACTTATCCGTATGAGAAAAACGCTTGTCTTTTACCGAGGTCGAGCGCCTAATG
GCCAAAAGTCCGATTGGATCATTCACGAATACTACAGCCTCGAGTCACACCAGAACTCTC

CTCCACAGGAAGAAGGATGGGTAGTGTGTAGAGCATTTAAGAAACGAACGACCATCCCAA
CAAAAAGGAGGCAACTTTGGGATCCGAACCTGCTTATTCTACGACGACGCCACTCTCTTGG
AACCTCTCGACAAGCGAGCCAGACATAATCCTGATTTTACCGCCACACCGTTCAAGCAAG
AACTACTCTCCGAGGCCAGTCACGTCCAGGATGGAGATTTTCGGATCTATGTACCTTCAAT
GCATCGATGATGATCAATTCTCCAGCTTCTCAGCTCGAGAGCCCCCTCTCTTCCGTCGG
AAATAACTCCCCATAGTACTACTTTTTCTGAGAACAGTAGCCGGAAGATGACATGAGCT
CCGAGAAGAGGATCACTGACTGGAGATATCTAGATAAGTTTCGTGGCGTCTCAATTTTGA
TGAGTGGAGAAGACTAAAAAAGGCTTTCCTATGCATGCATGCACTAGAAACGTCGTGCGA
TTTTGGATTACATGCGGCCGCT

>G1453 Amino Acid Sequence (conserved domain in AA coordinates:13-160)

MMKVDQDYSCSIPPGFRFHPTDEELVGYYLKKKIASQRIDLDVIREIDLYKIEPWLQER
CRIGYEEQTEWYFFSHRDKKYPGTGRTNRATVAGFWKATGRDKAVYLNLSKLIGMRKTLVF
YRGRAPNGQKSDWIIEHYYSLESHQNSPPQEKGWVVCRAFKKRTTIPTKRRQLWDPNCLF
YDDATLLEPLDKRARHNPFTATPFKQELLSEASHVQDGFSGMYLQCIDDDQFSQLPQL
ESPSPSEITPHSTTFSENSSRKDDMSSEKRITDWRYLDKFVASQFLMSGED*

>G1560 (120..1340)

ATCCTTTCAATTTCCACTCCTCTCTAATATAATTCACATTTTCCCCTATTGCTGATTCA
TTTTTTTTTGTGAATTATTTCAAACCCACATAAAAAATCTTTGTTTAAATTAAAAACCA
TGGATCCTTCATTTAGGTTTCATTAAGAGGAGTTTCTGCTGGATTTCAGTGATTCTCCAT
CACCACCATCTTCTTCTTACATCTTTATTCATCTTCCATGGCTGAAGCAGCCATAAATG
ATCCAACAACATTGAGCTATCCACAACCATAGAAGGTCTCCATGAATCAGGGCCACCTC
CATTTTTGACAAAGACATATGACTTGGTGGAAGATTCAAGAACCAATCATGTCTGTCTT
GGAGCAAATCCAATAACAGCTTTCATTGTCTGGGATCCACAGGCCTTTTCTGTAACCTCTCC
TTCCAGATTCTTCAAGCACAATAACTTCTCCAGTTTGTCCGCCAGCTCAACACATATG
GTTTCAGAAAGGTGAATCCGGATCGGTGGGAGTTTGCAAACGAAGGGTTTCTTAGAGGGC
AAAAGCATCTCCTCAAGAACATAAGGAGAAGAAAAACAAGTAATAATAGTAATCAAATGC
AACAACCTCAAAGTTCTGAACAACAATCTCTAGACAATTTTTGCATAGAAGTGGGTAGGT
ACGGTCTAGATGGAGAGATGGACAGCCTAAGGCGAGACAAGCAAGTGTGATGATGGAGC
TAGTGAGACTAAGACAGCAACAACAAAGCACCAAAATGTATCTCACATTGATTGAAGAGA
AGCTCAAGAAGACCGAGTCAAAAACAAAAACAATGATGAGCTTCTTGCCTCGCGCAATGC
AGAATCCAGATTTTATTTCAGCAGCTAGTAGAGCAGAAGGAAAAGAGGAAAGAGATCGAAG
AGGCGATCAGCAAGAAGAGACAAGACCGATCGATCAAGGAAAAGAAATGTGGAAGATT
ATGGTGATGAAAGTGTTATGGGAATGATGTTGCAGCCTCATCCTCAGCATTGATTGGTA
TGAGTCAGGAATATACATATGGAAACATGTCTGAATTCGAGATGTCCGAGTTGACAAAC
TTGCTATGCACATTCAAGGACTTGGAGATAATTCCAGTGCTAGGGAAGAAGTCTTGAATG
TGGAAAAAGGAAATGATGAGGAAGAAGTAGAAGATCAACAACAAGGTACCATAAGGAGA
ACAATTGAGATTATGTTGAAGTTTGTGGGAAGATTGTTAAATGAAGGTCAAAATTTTG
ATTTTGAAGGAGATCAAGAAATGTTGATGTGTTAATTCAGCAACTTGGTTATTGCGGTT
CTAGTTCACACACTAATTAAGAAGAAATGAAATGATGACTACTTTAAGCATTGGAATCA
ACTTGTTCCTATTAGTAATTTGGCTTTGTTTCAATCAAGTGAGTCGTGGACTAACTTGC
>G1560 Amino Acid Sequence (domain in AA coordinates: 62-151)

MDPSFRFIKEFFPAGFSDSPSPSSSSYLYSSSMAEAAINDPTTSLYPQPLEGLHESGPP
PFLT KTYDLVEDSRTNHVSVWSKSNNSFIVWDPQAFSVTL LPRFFKHNNFSSFVRQLNTY
GFRKVNPD RWEFANEGFLRGQKHL LKNIRRRKTSNNSNQMQPQSSEQQSLDNFCIEVGR
YGLDGEMDSLRRDKQVLMELVRLRQQQSTKMYLTLIEEKLKKTESKQKQMMSFLARAM
QNPDFIQQLVEQKEKRKEIEEAI SKKRQRPIDQGRNVEDYGDSEGYGNDVAASSSALIG
MSQEYTYGNMSEFEMSELDKLA MHIQGLGDNSSAREEVLNVEKGNDEEEVEDQQQGYHKE
NNEIYGE GFWE D L L N E G Q N F D F E G D Q E N V D V L I Q Q L G Y L G S S S H T N *

>G1594 (1..984)

ATGGATGGAATGTACAATTTCCATTTCGCCCGGTGATTATTCAGATAAGTCGGTTCTGATG
ATGTCACCGGAGAGTCTCATGTTTCCCTTCCGATTACCAAGCTTTGCTATGTTCTCCGCC
GGTGAATTCGTGTCTCTGATGTTTTCGGATCCGACGAGCTACTCTCAGTAGCCGTCTCC
GCTTTGTCGTCCGAGGCGGCTTCGATCGCTCCGAGATCCGAAGAAATGATGATAACGTT
TCTTAACGTGTCATCAAGCTAAATCGCTTGTCATCCTTCGTATCCTCGCTTACTTCAA
GCTTACATCGATTGCCAAAAGGTCCGAGCACCACCGAGATAGCGTGTCTTACTAGAGGAG
ATTCAACGGGAGAGTGATGTTTATAAGCAAGAGGTTGTTCTTCTTCTGCTTTGGAGCT
GATCCTGAGCTTGATGAATTTATGGAACGTACTGCGATATATTAGTGAAATACAAATCG

GATCTAGCAAGACCGTTTGTACGAGGCAACGTGTTTCTTGAACAAGATTGAGATGCAGCTA
CGGAACCTATGTACTGGTGTCTGAGTCTGCCAGGGGAGTTTCTGAGGATGGTGTAAATATCA
TCTGACGAGGAACTGAGTGGAGGTGATCATGAGGTAGCAGAGATGGGAGACAAAGATGT
GAAGACCGGGACCTCAAAGATAGGTTGCTACGCAAATTTGGAAGCCGTATTAGTACTTTA
AAGCTTGAGTTCTCAAAGAAGAAGAAGAAAGGAAAGTTACCAAGAGAAGCAAGACAAGCT
CTTCTTGATTGGTGAATCTCCATTATAAGTGGCCTTACCCTACTGAAGGAGATAAGATA
GCATTAGCTGATGCAACGGGGTTAGACCACAAAACAATCAACAATTGGTTTATAAACCAA
AGGAAACGTCAATTGGAAGCCATCAGAGAATATGCCTTTTCGCTATGATGGATGATTCTAGT
GGATCATTCTTTACCGAGGAATGA

>G1594 Amino Acid Sequence (conserved domain in AA coordinates:343-308)

MDGMYNFHSAGDYSDKSVLMMSPELMPSPDYQALLCSSAGENRVSDVFGSDELLSVAVS
ALSSEAAASIAPEIRNDDNVSLTVIKAKIACHPSYPRLLQAYIDCQKVGAPPEIACLLEE
IQRESDVYKQEVVPSSCFGADPELDEFMETYCDILVKYKSDLARPFDEATCFLNKIEML
RNLCTGVESARGVSEDGVISSDEELSGGDHEVAEDGRQRCEDRDLKDRLLRKFGSRISTL
KLEFSKKKKKGLPREARQALLDWWNLHYKWPYPTEGDKIALADATGLDQKQINNWFINQ
RKRHWKPSENMPFAMDDSSGSFFTEE*

>G1750 (94..1101)

CCCTTTTCTCTCTTTCTCCAAATCTCTGAAAATTTTACCAGAATCTCTGTTCTTTTTT
TCACCAGAATCTCTCTGTTTAAATAATAGGTGATGATGATGGATGAGTTTATGGATCTT
AGACCAGTGAAGTACACAGAGCACAAGACTGTTATCAGAAAGTACACTAAAAAGTCGTCT
ATGGAGAGGAAGACAGTGTTCGTGACTCGGCCAGGTTGGTTCCGGTCTCAATGACGGAT
CGTGACGCCACTGATTCATCAAGCGACGAGGAAGAGTTTCTGTTCCCTCGAAGACGTGTC
AAGAGATTGATTAACGAGATCAGAGTCGAGCCTAGCAGCTCTTCCACCGGCGACGTCTCT
GCTTCTCCGACGAAGGACCGGAAAAGAATCAACGTTGATTCTACGGTTCAAAGCCCTCT
GTTTCCGGCCAAAACCAAGAGAAGTACCGCGCGGTGAGACAGCGACCATGGGGAAAATGG
GCGGCGGAGATTCTGTGATCTGAGCAACGCCGAGAATCTGGCTCGGTACTTTTGCAACG
GCGGAGGAAGCTGCCATCGTCTACGACAACGCAGCAATCAAACCTCGTGGCCCTGATGCT
CTTACCAACTTCACCGTACAACCAGAACCAGAACCAGGTACAAGAACAAGAACAAGAACCG
GAGAGCAACATGTCGGTTTCGATATCAGAATCAATGGACGATTCTCAACATCTATCATCT
CCGACATCGGTTCTCAACTACCAAACATATGTCTCGGAGGAACCAATCGATAGTCTTATC
AAACCGGTTAAACAAGAGTTTCTTGAACCAGAACAGAGCCAATAAGCTGGCATCTTGGA
GAAGGTAATACTAATACTAATGATGATTCAATTTCCATTGGACATTACATTTCTCGACAAC
TATTTCAATGAATCATTACGACATCTCCATCTTCGATCAACCTATGTCTCCTATTCAA
CCAAACAGAGAATGATTTCTTCAACGACCTTATGTTATTTCGATAGCAACGCAGAAGAATAC
TACTCCTCCGAGATCAAAGAGATTGGTTTCATCGTTCAACGATCTTGATGATTCTTTGATA
TCCGATCTCTTACTGTGTGATATTTTGGCATTAAACCAAACACCGGTTTGGTTGC

>G1750 Amino Acid Sequence (domain in AA coordinates: 107-173)

MMMDDEFMDLRPVKYTEHKTVIRKYTKKSSMERKTSVRDSARLVRVSMTRDATDSSSDEE
EFLFPRRRVKRLINERVEPSSSSTGDVSASPTKDRKRINVDSTVQKPSVSGQNQKKYRG
VRQRPWGKWAEEIRDPEQRRIWLGTFAAEAAIIVYDNAAIKLRGPDALTNFTVQPEPE
PVQEQEQEPESNMSVSISESMDDSQLHSSPTSVLNYQTYVSEEPIDSLIKPVKQEFLEPE
QEPISWHLGEGNTNTNDDSFPLDITFLDNYFNESLPDISIFDQPMSPIQPTENDFFNDLM
LFDSNAEEYYSSEIKEIGSSFNDLDDSLISDLLLV*

>G1947 (70..918)

ACAACTATTCTCTCTCTCTCTTTTTTTTATTAAAAAGCTCAAATTTATATAGGTTTTTT
GTTCAAAAATGGATTATAACCTTCCAATTCATTAGAGGTCTCAAAGAAACGCCACCA
ACGGCTTTCTTGACGAAAACATACAACATAGTGGAGGATTCAAGCACAAACAACATAGTT
TCATGGAGCAGAGACAAACAGCTTCATTGTTTGGGAACCAGAGACTTTTGCCCTAATT
TGCCCTCCCTAGATGCTTTAAGCACAAATAATTTCTCCAGCTTTGTTAGACAGCTCAATACT
TATGGGTTTAAAGAGATTGATACAGAGAGATGGGAATTTGCAAATGAGCATTTTCTGAAG
GGAGAGAGGCATCTTCTTAAGAACATCAAGAGAAGAAAGACATCTCAAACGCAAACG
CAGTCGCTAGAAGGAGAGATCCATGAGCTGCGAAGAGACAGAATGGCTTTAGAAGTAGAA
CTGGTTAGACTGCGACGAAAACAAGAAAGCGTGAAGACATATCTGCATTTGATGGAAGAG
AAACTGAAAGTCACAGAAGTAAAGCAAGAAATGATGATGAATTTCTTGCTAAAGAAGATT
AAGAAACCGAGTTTTTTTACAGAGCTTAAGGAAACGTAATCTGCAAGGAATCAAGAATCGA
GAGCAAAAGCAAGAGGTGATCTCAAGCCATGGTGTGAGGATAATGAAAAGTTTGTAA
GCTGAGCCAGAAGAGTATGGTGATGACATCGATGATCAATGTGGAGGTGTGTTTGATTAT

GGTGATGAGCTTCACATAGCTTCAATGGAGCATCAAGGACAAGGGGAGGATGAAATTGAA
ATGGATAGTGAAGGAATTTGGAAGGGTTTCGTGTGAGTGAGGAGGAGATGTGTGATTGA
GTGGAACATTTTATATAATAAACTAATGTATTATGAGAGGTTTTTTTTGTTTTTTTGCT
TTTTTTTTCCGAGTTTGTCAATGATGATCAAGCATTGTATACAATTTGGGCCAACTAAAAGCCCAA
CAAAATATTTGGCCTTGGCATTGTGTAACAAATTGACTAATTCGGCCACACCTTCC
>G1947 Amino Acid Sequence (domain in AA coordinates: 37-120)
MDYNLPIPLEGLKETPPTAFLTKTYNIVEDSSTNNIVSWSRDNNSFIVWEPETFALICLP
RCFKHNNFSSFVRQLNTYGFKKIDTERWEFANEHFLKGERHLLKNIKRRKTSSQTQTQSL
EGEIHLELRDRMALEVELVRLRRKQESVKTYLHLMEEKLVTEVKQEMMNFLKKIKKP
SFLQSLRKRNLQGIKNREQKQEVISSHGVEDNGKFVKAEPPEYDGDIDQCGGVFDYGD
LHIASMEHQGGEGEIEIEMDSEGIWKGFLVSEEMCDLVEHFI*
>G2011 (309..1547)
AATGTCGGTTGTACAATATTTGTCTACTAAAGTTTCCAAATTTCTTCTAAACTGATGAAT
CAATGGAACATGATGACGAAAAAGATAAATCCACGGTGGCGGGAAGTACCCACCCATTT
CCACCGCCTCTCTATTTCCCGAGATTTTTTCAATTATCTGACTACAGTTTGTCTGTTACT
TCCTTCCCTAAACCTTTATAAACCATTAACCTCTCATCTTCTTCTTAAACCCCTTA
ATTATCACACACCCCAATTTCTCACTCTCTCTCTCACTAAACCCGTAATTTTCTAC
TATATCAAATGAGCCAAAAAGATGCTGTTTCTAAACCAACTCCAATTTTCAGTACCCG
TTTCGAGACGATCCGATATACCCGGTCTCTCTACGTCGACACTGACATGGGTTTCTCTG
GGTCACCACTTCCCATGCCACTAGACATCTTACAAGGGAATCCAATTTCCACCTTTTTAT
CCAAGACTTTTGATTTGGTTGATGACCCGACTCTTGACCCGGTCATCTCTTGGGGACTGA
CCGGAGCTAGCTTCGTAGTTTGGGATCCTCTAGAGTTTGGCAGAATCATACTTCCAAGGA
ATTTCAAACACAACATTTCTCCAGCTTCGTCAGACAGCTTAACACTTATGGATTTGAA
AGATTGATACTGACAAGTGGGAATTCGCTAACGAGGCTTTCTTAGAGGCAAGAAGCATC
TTCTGAAGAACATTATCGTCGTCGATCACCACAATCCAACCAAACTTGCTGCAGTAGCA
CTAGCCAAAGCCAAGGGTCACCTACTGAGGTTGGAGGAGAGATTGAGAAGCTGAGGAAAG
AGCGGCGTGCAATGATGGAGGAAATGCTTGAGCTTCAAGCAGCAAGCAGAGGCACAGCTC
GACATGTGGACACTGTAAACCAGAGGCTGAAAGCTGCAGAGCAACGTCAGAAAGCAATTGC
TCTCTTTCTTGGCTAAGTTGTTTCAAGAACCGGGTTTCTTGGAAACGCTGAAGAACTTCA
AAGGAAAAGAAAAGGAGGAGCTTGGATTGGAAGGCGAGAAAGAGTTTCAATCAAGC
ACCACCAGCAGCTCAAGATTCTCAACAGGAGGGGAGGTGGTGAAGTATGAAGCTGATG
ATTGGGAGAGATTGCTAATGTATGACGAAGAGACTGAGAACACCAAGGGTTTAGGAGGGA
TGACTTCAAGCGATCCAAAGGCAAGAACTTGATGTATCCATCAGAAGAAGAGATGAGCA
AACCAGATTACTGATGTCTTCCCATCTCTGAAGGACTTATTAACAAGAAGAGACGA
CATGGAGCATGGGTTTCGATACACAGTTCAGCAACACCGATGCATGGGGAA
ACACAATGGACTATAATGATGTCTCAGAGTTTGGTTTGTCTGCAGAAACAACAAGTATG
GTTTGCCTGATGTCTGCTGGGAACAATTTGCTGCAGGAATCAAGAGACTGGATTCAACT
GGCCAACCTGGTATGATGATGATAATACGCCAATGAATGATCCTTAGGATCTTTTCATAT
ATAGTTTACACCAAAAACCGTTTCTTATCGGGTGAATATTAATTCATTATTCTTTTG
AATGCACTCTTTATACATATATATAATATTGATGAGTTTGATTGTTCCAAAAA
>G2011 Amino Acid Sequence (domain in AA coordinates: 56-147)
MSPKKDAVSKPTPIVPSVRSRSDIPGSLYVDITDMGFSGLPMPDLILQGNPIPPFLSKT
FDLVDDPTLDPVISWGLTGASFVWDPLEFARIILPRNFKHNNFSSFVRQLNTYGFRIKID
TDKWEFANEAFRLRGKKHLLKNIHRRRSPQSNQTCSSSTSQSQSGSPTEVGGEIEKLRKERR
ALMEEMVELQQQSRGTARHVDVTNQRLLKAAEQRQKQLLSFLAKLFQNRGFLERLKNFKGK
EKGGALGLEKARKKFIKHQQPQDSPTGGEVVKYEADDWERLLMYDEETENTKGLGGMTS
SDPKGKNLMPSEEEEMSKPDYLSFSPSPEGLIKQETTWSMGFDTTIPSFSNTDAWGNTM
DYNDVSEFGFAAETTS DGLPDVCWEQFAAGITETGFNWPTGDDDDNTPMNDP*
>G2094 (1..450)
ATGCTAGATCCCACGAGAAAGTAATCGATTGAGAAATCAATGGAAAGCAAACCTCACATCA
GTAGATGCGATCGAAGAACACAGCAGCAGTAGCAGTAATGAAGCTATCAGCAACGAGAAG
AAGAGTTGTGCCATTTGTGTACAGCAAAACCCCTCTTTGGCGAGGCGGTCTTGCCGGT
CCCAAGTCGCTTTGTAACGCATGCGGGATCAGAAACAGAAAGAAAGAAAGAACTGATC
TCAAATAGATCAGAAGATAAGAAGAAGAGTCAACAGAAACCCGAAGTTTGGTGAC
TCGTTGAAGCAGCGATTAAATGGAATTGGGGAGAGAAGTGTATGATGCAGCGATCAACGGCT
GAGAATCAACGGCGGAATAAGCTTGGCGAAGAAGAGCAAGCCGCGTGTACTCATGGCT
CTCTCTTATGCTTCTCCGTTTATGCTTAA

>G2094 Amino Acid Sequence (domain in AA coordinates:43-68)
MLDPTEKVIDSESMESKLTSDVAIEEHSSSSSNEAISNEKKSCAICGTSKTPLRGGPAG
PKSLCNACGIRNRKRRTLISNRSEDKKKKSHNRNPKFGDSLKQRLMELGREVMMQRSTA
ENQRRNKLGEEEQAAVLLMALSYASSVYA*

>G2113 (90..590)
ATAACAACTCATCAAACCTCCTCAGCGTTTCTTTTCTTACATAAACAATTTTCTTAC
ATAAACAAATCTTGTTGTTGTTGTCATGGCACCGACAGTTAAACCGCGGCCGTCA
AAACCAACGAAGGTAACGGAGTCCGTTACAGAGGAGTGAGGAAGAGACCATGGGGACGTT
ACGCAGCCGAGATCAGAGATCCTTTCAAGAAGTCACGTGTCTGGCTCGGTACTTTGACA
CTCCTGAAGAAGCCGCTCGTGCCTACGACAAACGTGCTATTGAGTTTCTGAGCTAAAG
CCAAAACCAACTTCCCTTGTTACAACATCAACGCCCACTGCTTGAGTTTGACACAGAGCC
TGAGCCAGAGCAGCACCGTGGAATCATCGTTTCTTAATCTCAACCTCGGATCTGACTCTG
TTAGTTCGAGATTCCCTTTTCTTAAGATTGAGTTAAGGCTGGGATGATGGTGTTCGATG
AAAGGAGTGAATCGGATTCTTCGTGCGTGGTGGATGTCGTTAGATATGAAGGACGAC
GTGTGGTTTGGACTTGGATCTTAATTTCCCTCCTCCACCTGAGAACTGATTAAGATTTA
ATTATGATTATTAGATATAAATTAAATGTTTCTGAATTGAG

>G2113 Amino Acid Sequence (domain in AA coordinates: TBD)
MAPTVKTAAVKTNEGNGVRYRGVVRKRPWGRYAAEIRDPFKSRVWLGTFDTPEEAARAYD
KRAIEFRGAKAKTNFPCYNINAHCLSLTQSLQSSTVESSFPNLNLGSDSVSSRFPFPI
QVKAGMMVFDERSESDSSSVMDVVRYEGRRVVLDDLNLNFPFPPEN*

>G2115 (41..733)
AATCACTCTACAAAGCCTGTACGTACACAACAACATTACCATGGTGAAACAAGAACGCAA
GATCCAAACCAGCAGCACAAAAAGGAAATGCCTTTGTCTCATCATCACCATCTTCTTCTC
TTCTTTCATCTTCTTCTTCTTCTGTCCTTTCGTGTAAGAACAAGAACAAGAAGAGTAAGAT
TAAGAAGTACAAAGGAGTGAGGATGAGAAGTTGGGGATCATGGGTCTCTGAGATTAGGGC
ACCAAATCAAAAGACAAGGATTGTTAGGTTCTTACTCAACAGCTGAAGCAGCTGCTAG
AGCTTACGATGTTGCACTCTTATGTCTCAAAGGCCCTCAAGCCAATCTCAACTTCCCTAC
TTCTTCTTCTTCTCATCATCTTCTTGATAATCTCTTAGATGAAAATACCCCTTTGTCCCC
CAAATCCATCCAAAGAGTAGCTGCTCAAGCTGCCAATCATTTAACCATTTTGCCCTAC
TTCATCAGCCGTCCTCGTCACCGTCCGATCATGATCATCACCATGATGATGGGATGCAATC
TTTGATGGGATCTTTTGTGGACAATCATGTGTCTTTGATGGATTCAACATCTTCATGGTA
TGATGATCATAATGGGATGTTCTTGTGATAATGGAGCTCCATTCAATTACTCTCCTCA
ACTAAACTCGACGACGATGCTCGATGAATACTTCTACGAAGATGCTGACATTCCGCTTTG
GAGTTTCAATTAATCCGACGGTCCATAATACATACTTTAATTAGT

>G2115 Amino Acid Sequence (conserved domain in AA coordinates:46-115)
MVKQERKIQTSSTKKEMPLSSSSSSSSSSSSSSSSSSCKNKNKSKIKKYKVRMRSWGS
WVSEIRAPNQKTRIWLGSYSTAEAAARAYDVALLCLKGPQANLNFTSSSSHLLDNLDD
ENTLLSPKSIQRVAAQAANSFNHFAPTSSAVSSPSDHDHHDHDDGMQSLMGSFVDNHVSLM
DSTSSWYDDHNGMFLFDNGAPFNYSPLNSTTMLDEYFYEDADIPLWSFN*

>G2130 (41..988)
CCTCTCTTCATTTTAACTCCCTCTCTCTCTCTCTCTATGGAGAGACGAACGAGACG
AGTGAAGTTCACAGAGAATCGTACGGTCACAAACGTAGCAGCTACACCATCTAACGGGTC
TCCGAGACTGGTCCGTATCACTGTTACTGATCCTTTTCGCTACTGACTCGTCTAGCGACGA
CGACGACAACAACACGTACCGTGGTTCCAAGAGTGAACGATACGTGAAGGAGATTAG
ATTCTGCCAAGGTGAATCTTCTTCTCCACCGCGGCGAGGAAAGGTAAGCACAAAGGAGGA
GGAAAGCGTAGTGGTTGAAGATGACGTGTGACGTGCGTGAAGCCTAAAAAGTACAGAGG
CGTGAGACAGAGACCTTGGGGAAAATTGCGCGCGGAGATTAGAGATCCGTCGAGCCGTAC
TCGGATTTGGCTTGGGACTTTTGTACGGCGGAGGAAGCTGCTATAGCGTACGATAGAGC
CGCGATTCTCTCAAAGGACCTAAAGCGCTCACGAATTTCTTAATCTCCGCCGACGCCAAC
GCCGGTTATCGATCTCCAAACGGTTTCCGCTGCGATTACGGTAGAGATTCTCGGCAGAG
CCTTCATTACCCGACCTCTGTTCTAAGATTCAACGTCAACGAGGAAAACAGAGCATGAGAT
TGAAGCGATCGAGCTATCTCCGAGAGAAAGTGCACGGTTATAAAAGAAGAAGAAGATC
GTCGGCGGGTTTGGTGTTCCTGGATCCGTATCTGTTACCGGATTTATCTCTCGCCGGCGA
ATGTTTTTGGGATACCGAAATTGCCCTGACCTTTTGTCTCGATGAAGAAACCAAAT
CCAATCAACGTTGTTACCAAAACACAGAGGTTTCGAAACAAGGAGAAAACGAAACTGAAGA
TTTCGAGTTTGGTTGATTGATGATTTTCGAGTCTTCTCCATGGGATGTGGATCATTTCTT
CGACCATCATCACTCTTTCGATTAAAAATCTCTTCTTTTGGGGAAATTTTGTG

>G2130 Amino Acid Sequence (domain in AA coordinates 93-160)
MERRTRRVKFTENRTVTNVAATPSNGSPRLVLRITVTDPFATDSSSDDDNNNVTVVPRVK
RYVKEIRFCQGESSSSSTAARKGKHKEESVVVEDDVSTSVKPKKYRGVRQRPWGKFAAEI
RDPSSRTRIWLGTFTVTAEEAAIAYDRAAIHLKGPALTNFLTPTPTPVIDLQTVSACDY
GRDSRQSLHSPTSVLRFNVNEETEHEIEAIELSPERKSTVIKEEEESSAGLVFPDPYLLP
DLSLAGECFWDTEIAPDLLFLDEETKIQSTLLPNTEVSKQGENETEDFEFGLIDDFESSP
WDVDHFFDHHHHSFD*

>G2147 (162..1262)

CTGTGATTGTCAAGAGTTTGAACACACAAAGAAGAAAGAACTCAACATTTCAAGCAA
GAAGAAAGAGAGAAGAGAGAAGGTCCAATAATAGAGAGAACAAAAAAGAGAGCTTAA
TTGTCTAGTTTATTCTCTGCAACCGTGCAGGCTTAAGTAACACATGTCTGAATTTATGGAGTTA
AAGAGCTCACATGGGAAAATGGGCAACTAACCGTTCATGGTCTAGGCGACGAAGTAGAAC
CAACCACCTCGAATAACCCCTATTGGACTCAAAGTCTCAACGGTTGTGAGACTTTGGAGT
CTGTGGTTCATCAAGCGGCTCTACAGCAGCCAAGCAAGTTTCAGCTGCAGAGTCCGAATG
GTCCAAACCACAATTATGAGAGCAAGGATGGATCTTGTTCAGAAAACCGGGTTATCCTC
AAGAAATGGACCGATGGTTCGCTGTTCAGAGGAGAGCCATAGAGTTGGCCACAGCGTCA
CTGCAAGTCCGAGTGGTACCAATATGTCTTGGGCGTCTTTTGAATCCGGTCCGAGCTTGA
AGACAGCTAGAACCGGAGACAGAGACTATTTCCGCTCTGGATCGGAACTCAAGATACTG
AAGGAGATGAACAAGAGACAAGAGGAGAAGCAGGTAGATCTAATGGACGACGGGGACGAG
CAGCAGCGATTACAACGAGTCCGAAAGGAGACGGCGTGATAGGATAAACCAGAGGATGA
GAACACTTCAGAAGCTGCTTCTCTACTGCAAGTAAGGCGGATAAAGTCTCAATCTTGGATG
ATGTTATCGAACACTTGAACAGCTACAAGCACAAAGTACAGTTTCATGAGCCTAAGAGCCA
ACTTGCCACAACAATGATGATTCCGCAACTACCTCCACCACAGTCAGTTCTCAGCATCC
AACACCAACAACAACAACAACAGCAGCAGCAGCAGCAACAACAGCAGCAACAGTTTC
AGATGTCGTTGCTTGCAACAATGGCAAGAATGGGAATGGGAGGTGGTGGAAATGGTTATG
GAGGTTTAGTTCTCTCTCTCTCTCTCCACCAATGATGGTCCCTCTCTATGGGTAACAGAG
ACTGCACCAACGGTCTCTTCAGCCACATTATCTGATCCATACAGCGCTTTTTCGCACAGA
CAATGAATATGGATCTCTACAATAAATGGCAGCAGCTATCTATAGACAACAGTCTGATC
AAACAACAAGGTAAATATCGGCATGCCTTCAAGTTCTTCAAGTCTGAGAAAAGAGATT
AGTCTAGCGACCTAGTATTATTGATCCATATATAGTTCTTGAAAGATTGTTGTATCAT
GATTGTAATAAAGTGTGTTGAGTATGGAAGAAAGACTTGACAGATAAAA

>G2147 Amino Acid Sequence (domain in AA coordinates:160-234)
MSNYGVKELTWENGQLTVHGLGDEVEPTTSNNPIWTQSLNGCETLESVVHQAALQQPSKF
QLQSPNGPNHNYESKDGSCSRKRGYPOEMDRWFAVQEEESHVRVGHSVTASASGTNMSWASF
ESGRSLKTARTGDRDYFRSGSETQDTEGDEQETRGEAGRSNGRRGAAAAIHNESERRRRD
RINQRMRTLQKLLPTASKADKVSILDDVIEHLKQLQAQVQFMSLRANLPQQMMIPQLPPP
QSVLSIQHQQQQQQQQQQQQQQQQFQMSLLATMARMGMGGGNGYGGGLVPPPPPPPMV
PPMGNRDCTNGSSATLSDPYSAFFAQTMNMDLYNKMAAAIYRQSDQTTKVNIGMPSSSS
NHEKRD*

>G2156 (384..1292)

TTTTTTTCCCTTTCTCGTTCAAAAAAGTACTTGCAGAGTCACTCACTCTCAGTCTCA
GCACATGAATTAATTTGAAGCTTCCCTAGAATTCTTTACATCAATTAATACGACACCGT
CTCGGTGAAGAATCTCTCTCTCTTGGCCCTAAAGCGAGTTAGGGTTTAACACACAAAGC
ATACCCTTTAGATTGTGTCTCTTAGCTCTGTTTTGTGCGGCTTGTGTAACCGATCAACT
CAAGCTATTGGCTCTCACCTCCTGAAATTTGACTTCTCCAATGGATCTCAAAGTTTCTC
TTATATGAATTCTATCTTCACTTCAATATCTTTATATATATGAGCCACAAGAACAAG
AAGAGTCAGTAGATGCGGCTGCCATGGACGGTGGTTACGATCAATCCGGAGGAGCTTCTA
GATACTTTCACAACCTCTTCAGGCCTGAGCTTCATCACCAGCTTCAACCTCAGCCTCAAC
TTCACCTTTGCTCAGCCTCAGCCTCAACCTCAGCCTCAGCAGCAGAATTCAGATGATG
AATCTGACTCCAACAAGGATCCGGGTTCCGACCCAGTTACCTCTGGTTCAACCGGGAAC
GTCCACGTGGACGTCTCCGGGATCCAGAACAAGCCGAAGCCACCGGTGATAGTACTA
GAGATAGCCCCAACGTGCTTAGATCTCATGTTCTTGAAGTCTCATCTGGAGCCGACATAG
TCGAGAGCGTTACCACTTACGCTCGCAGGAGAGGAAGAGGAGTCTCCATTCTCAGTGGTA
ACGGCACGGTGGCTAACGTGAGTCTCCGGCAGCCGGCAACGACAGCGGCTCATGGGGCAA
ATGGTGAACCGGAGGTGTTGTGGCTCTACATGGAAGGTTTGAGATACTTTCCCTCACAG
GTACGGTGTGCGGCCCCCTGCGCCGCCAGGATCCGGTGGTCTTTCTATCTTTCTTCCG
GCTTCAAGGTCAGGTGATTGGAGGAAACGTGGTGGCTCCGCTTGTGGCTTCGGGTCCAG

AACGGTGAAGGCCGATGGAATCTGCTCGCTAAATCTTCTGGGCTAAAGAGAGCAGGAAAA
AGTTGTAGATTGAGATGGTTGAATTACCTTAAACCCGACATAAAGCGTGGGAATCTCACT
CCTCAAGAACAACCTTTAATCCTTGAGCTCCATTCTAAATGGGGTAATAGGTGGTCAAAA
ATTTCAAGTATTTACCAGGAAGAACAGACAACGATATCAAAAACCTACTGGAGAACTAGA
GTCCAGAAACAAGCAGCCAGCTCAACATAGATTCCAATAGCCACAAGTTCATAGAAGTT
GTTCTAGCTTTTGGTTTCCAAGACTGATCAACGAGATTAAAGACAACCTCATAACCAAC
AATATTAAGCTAATGCTCCTGATTTACTTGGACCAATTTTACGAGACAGCAAAGATTG
GGTTTCAACAACATGGATTGTTCCACTTCCATGTCAGAAGATCTCAAGAAAACCTCACAA
TTCATGGATTTTCTGATCTTGAAACCACAATGTCCTTGAAGGATCACGAGGGGGTAGT
AGTCAATGTGTGAGTGAGGTTTATAGCTCCTTCCCTTGCCTAGAGGAGGAGTACATGGTG
GCCGTTATGGGCAGTTTCAGACATTTTCAGCATTCATGATTGTCACGTGGCTGATTTCAAG
TACGAGGATGATGTGACACAAGATCTAATGTGGAACATGGATGACATTTGGCAGTTTAAAC
GAGTATGCACACTTTAATTAGGTTATATTATTTATGTACTTCTTACAACCTGGAGGGG
TTTATCGGTCTTTTATTAAATTTGATTGTTTGGATTCTTTAAAAATGTGTTCTTATTA
TAGTTTTTAATGAAAAAATGTTTAAAGCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>G2893 Amino Acid Sequence (conserved domain in AA coordinates:19-120)
MSNITKKKCNNGEEGAEQRKGPWTLLEEDLLTNYISHNGEGRWNLLAKSSGLKRAGKSCR
LRWLNLYLKPDIKRNLTTPQEQLLILELHSHKGNRWSKISKYLPGRTDNDIKNYWRTRVQK
QARQLNIDSNSHKFIEVVRSFWFPRLINEIKDINSYNNIKANAPDLLGPILRDSKDLGFN
NMDCSTSMSEDLKKTQSFMDFSDLTTMSLEGRSGSSQCVSEYSSFPCLSEYEMVAVM
GSSDISALHDCHVADSKYEDDVTQDLMWNMDDIWQFNEYAHFN*
>G340 (97..834)
ATGAAATCTCTGTAGTTTTTTTTTGTTCCTTTCTTAAATTTTCGAAAGAAAGACATTTATT
AAACCAAAATAACTCTTTAGATCATTGCAAGGAAAATGTTGAAAAGTGCAAGTCCAATG
GCATTCTACGATATCGGAGAGCAGCAATACTCTACTTTTCGGGTACATTTTAAGCAAACCT
GGGAACGCAGGAGCTTACGAGATTGACCTTCGATCCCAAACATCGACGATGCGATCTAC
GGCTCAGATGAGTTCCGTATGTACGCTTACAAAATCAAACGGTGTCTCTGACTCTGATGC
CACGACTGGACGGAGTGTCCCTACGCTCACCCTGGCGAGAAAGCCACACGCCGTGATCCT
CGCCGTTACACTTACTGTGCAGTCGATGCCCGGCTTTCCGAAATGGCGCATGCCACCGT
GGCGACTCATGCGAAATTCGCACATGGCGTATTCGAGTACTGGCTCCACCCGGCGGTTAC
CGAACACGCGCATGTAACGCCCGGGAACCTTGTGTGAGAGGAAAGTGTGTTCTTTGCCAC
GCGCCGGAGCAGCTAAGGCAGTCTGAAGGAAAGCACAGGTGCAGGTACGCATATAGGCCG
GTGAGGGCTAGAGGTGGTGAAACGGCGATGGAGTGACGATGAGAATGGACGACGAGGGT
TACGACACGTACCGTCTCCGGTGAGAAGCGGGAAGATGATTTAGATAGTAACGAGGAG
AAGTGTGTTGAGTGTGAGTGGAGTGGATGAGCATTGTGGATGATCATTATGAGCCGTCC
GATTTGGATTTGGATTTGTCACTTTGATTGGATCTCAGAGTTGGTCGATTAAATTTGG
GAAATCAAAGCAGAGAAACAAAAGAAACCCGATAAATAAAGTGGATTTTGTGTTAAATCCAC
AAGATCAAGATTCAAGATGAGAGATCTTGTCTATGTATATGGTAAATTTAATTGTAATGAT
TTATTGCAATGTCGCAAAAGAAGTTACTTCTCTTTCATGTAAACAGATTCTTGATCTTC
TATAAGTCTTTGTATTAA
>G340 Amino Acid Sequence (domain in AA coordinates: 37-154)
MLKSASPMFYDIGEQYSTFGYILSKPGNAGAYEIDPSIPNIDDAIYGSDEFMYAYKI
KRCPRTRSHDWTECPYAHRGEKATRRDPRRYTYCAVACPFRNGACHRGDSCEFAHGVFE
YWLHPARYRTRACNAGNLCQRKVCFFAHAPEQLRQSEGHKRCRYAYRPVRARGGGNGDGV
TMRMDDEGYDTSRSPVRSGKDDLSNEEKVLLKCSRMSIVDDHYEPSDLDDLDSHFDWI
SELVD*
>G39 (75..638)
GTTTCCACAGTCCCTGTACTTGTGCATAAACTGTAAACACTACTCTGAAAATTTTGCT
TCTGTTAGGATATAATGCCACCCTCTCTCCTAAATCTCCTTTTATTAGCTCTTCACTCA
AAGGAGCTCATGAAGATCGCAAATTTAAATGCTATAGGGGTGTCCGAAAGAGGTCTTGGG
GCAAATGGGTGCTGAAATCAGAGTTCCAAAGACTGGACGACGAATATGGCTAGGTTTCAT
ACGATGCTCCAGAGAAAGCAGCTAGAGCCTATGATGCTGCTTTGTTCTGTATTAGGGGTG
AGAAGGGAGTTTACAATTTTCCCACTGATAAAAAGCCGAGCTTCAGAAAGGTTCTGTCC
GGCCTCTGTCCAAGCTCGACATACAGACAATAGCAACAACTATGCTTCATCAGTTGTGC
ATGTACCTTCCCATGCCACCACACTCCCGGCAACAACCCAGGTTCCCTCTGAAGTTCCTG
CTTCCTCTGATGTTTCTGCTTCTACTGAGATTACAGAGATGGTCGATGAATATTATCTCC
CAACCGATGCAACTGCAGAATCAATATTCTCAGTTGAAGACTTACAACCTGGACAGTTTCC

TCATGATGGACATTGATTGGATAAACAATCTAATCTGATGTGTAACGTCACCTGCAGTGA
CATTTAATATGGTTTANCTATCAGTTACCTGTCTGCTTCTTGTAAAGGTATACTTGGATC
CTTGCTCTTGAACCTGTTTTATTAGCATGCAAA

>G39 Amino Acid Sequence (domain in AA coordinates: 24-90)

MPPSPKSPFISSSLKGAHEDRKFCKYRGVRKRSWGKVVSEIRVPKTGRRIRWLGSYDAPE
KAARAYDAALFCIRGEKGVNFPDCKPQLPEGSVRPLSKLDIQTATNYASSVVHVP
SH
ATTLPATTPVPSEVPASSDVSASTEITEMVDEYLLPTDATAESIFSVEDLQLDSFLMMDI
DWINNL*

>G439 (128..967)

TATAAATCTTCGTTTCTACTTTTTTTTCTTCCATAATATAGTCAATTCGTTTCTTAATT
AGGGCTTCTTCTCTTTGTTTCTCCAATCTTTATTAGTTTATTTATTTATTTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATTCATGAAGGTAACCTCATCTTCTTCTACTTCAATTCATCAAAATCCAAAACC
ATTAATCCTTAATTCATCCCTAATAATGACCAAGTCTTACCGGTATCTAACCACAAACCGG
TCCGATTGGGCTAAACCAGCTCACTCCAACACAAATCCTCCAAATTCAGACAGAGTTACA
TCTCCGGCAAAACCAATCTCGTCGTCGCGCTGGTAGTCATCTTCTACCGCTAAACCAAC
CTCAATGAAGAAAATCGACGTAGCAACTAAACCGGTTAAACTATAACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTGCGCTACCTAAAAACCGAACCCGGTTATG
GCTCGGTACGTTGCAACGGCTCAAGAAGCTGCATTAGCTTACGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTGTTGCTCAAGGACACTATAA
ACAGATATTGTCTCGTCTATCAACGCAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCTCAGATCGAGAAACAGAACAAAACAGAGGAGGTGCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCGGAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGACAAGTATCCTTCTTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTATCTTTTTTGATTTCATTTGTTCTCTAAATTGTAGAATTTTATTTTC
AGAGCTTTGTAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGGAACCTAACTCT
GTTTTCTTTTGTAAATGATTGTTTATAATGGGCCGTTGAATGGGCCTTATTGATTTAAACA
GCCCAAGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)

MAMALNMNAYVDEFMEALEPFMKVTSSTSSSNPKPLTPNFI PNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPVKLYRGVRQRQ
WGKWAIEIRLPKNRRLWLGTFFETAQEAALAYDQAHKIRGDNARLNFDPIDVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEBVLSGFSKPEKEPEFGEIYGCYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G470 (1..2580)

ATGGCGAGTTTCGAGGTTTCAATGAAAGGTAATCGTGGAGGAGATAACTTCTCCTCCTCT
GGTTTTAGTGACCCTAAGGAGACTAGAAATGTCTCCGTCGCCGCGGAGGGGCAAAAAGT
AATTCACCCGATCCGCTGCGGCTGAGCGTGCTTTGGACCTGAGGCTGCTCTTTACAGA
GAGCTATGGCAGCTTGTGCTGGTCCGCTTGTGACGGTTCCTAGACAAGACGACCGAGTC
TTCTATTTTCTCAAGGACACATCGAGCAGGTGGAGGCTTCGACGAACCGGCGGAGAA
CAACAGATGCCTCTCTATGATCTTCCGTCAAAGCTTCTGTCGAGTTATTAATGTAGAT
TTAAAGGCAGAGGCAGATACAGATGAAGTTTATGCGCAGATTACTTCTTCTCCTGAGGCT
AATCAAGACGAGAATGCAATGAGAAAGAAGCGCCTCTTCTCCACCTCCGAGGTTCCAG
GTGCATTGCTTCTGCAAAACCTTGACTGCATCCGACACAAGTACACATGGTGGATTTTCT
GTTCTTAGGCGACATGCGGATGAATGTCTCCACCTCTGGATATGTCTCGACAGCCTCCC
ACTCAAGAGTTAGTTGCAAAAGATTGTCATGCAAAATGAGTGGCGATTTCAGACATATATTC
CGGGGTCAACCACGAGGCAATTTGCTACAGAGTGGGTGGAGTGTGTTTGTAGCTCCAAA
AGGCTAGTTGCAGGCGATGCGTTTATATTTCTAAGGGGCGAGAATGGAGAATTAAGAGTT
GGTGTAAAGCGTGCGATGCGACAACAAGGAAACGTGCCGTCTTCTGTTATATCTAGCCAT
AGCATGCATCTTGAGTACTTGGCCACCGCATGGCATGCCATTTCACAGGGACTATGTTT
ACAGTCTACTACCAACCCAGGACGAGCCCATCTGAGTTATTGTTCCGTTTCGATCAGTAT
ATGGAGTCTGTTAAGAATACTACTCTATTGGCATGAGATTCAAAATGAGATTTGAAGGC
GAAGAGGCTCCTGAGCAGAGGTTTACTGGCACAATCGTTGGGATTGAAGAGTCTGATCCT
ACTAGGTGGCCAAAATCAAAGTGGAGATCCCTCAAGGTGAGATGGGATGAGACTTCTAGT
ATTCTCGACTGATAGAGTATCTCCGTGGAAAGTAGAGCCAGCTCTTGCTCCTCCTGCT
TTGAGTCCTGTTCCAATGCCTAGGCCTAAGAGGCCAGATCAAATATAGCACCTTCATCT

CCTGACTCTTCGATGCTTACCAGAGAAGGTACAACCTAAGGCAAACATGGACCCCTTTACCA
GCAAGCGGACTTTCAAGGGTCTTGCAAGGTCAAGAATACTCGACCTTGAGGACGAAACAT
ACTGAGAGTGTAGAGTGTGATGCTCCTGAGAATTCTGTTGTCTGGCAATCTTCAGCGGAT
GATGATAAGGTTGACGTGGTTTCGGGTTCTAGAAGATATGGATCTGAGAACTGGATGTCC
TCAGCCAGGCATGAACCTACTTACACAGATTTGCTCTCCGGCTTTGGGACTAACATAGAT
CCATCCCATTGGTCAGCGGATACCTTTTTATGACCATTTCATCATCACCTTCTATGCCTGCA
AAGAGAATCTTGAGTGATTGAGAAGGCAAGTTCGATTATCTTGCTAACCACTGGCAGATG
ATACACTCTGGTCTCTCCCTGAAGTTACATGAATCTCCTAAGGTACCTGCAGCAACTGAT
GCGTCTCTCCAAGGGCGATGCAATGTTAAATACAGCGAATATCCTGTTCTTAATGGTCTA
TCGACTGAGAATGCTGGTGGTAACTGGCCAATACGTCCACGTGCTTTGAATTATTATGAG
GAAGTGGTCAATGCTCAAGCGCAAGCTCAGGCTAGGGAGCAAGTAACAAAACAACCCCTC
ACGATACAAGAGGAGACAGCAAAGTCAAGAGAAGGGAAGTGCAGGCTCTTTGGCATTCTCT
CTGACCAACAACATGAATGGGACAGACTCAACCATGTCTCAGAGAAACAACCTGAATGAT
GCTGCGGGGCTTACACAGATAGCATCACCAAAGGTTCAAGACCTTTTCAAGATCAGTCAAAA
GGGTCAAAATCAACAAACGATCATCGTGAACAGGGAAGACCATTCCAGACTAATAATCCT
CATCCGAAGGATGCTCAACAGAAAACCAACTCAAGTAGGAGTTGCACAAAGGTTTCAAG
CAGGGAATTGCACTTGGCCGTTTCAAGTTCAGTGGATCTTTCAAAGTTCAAAACATATGAGGAGTTA
GTCGCTGAGCTGGACAGGCTGTTTGAAGTTCATGAGAGTTGATGGCTCCTAAGAAAGAT
TGGTTGATAGTTTACACAGATGAAGAGAATGATATGATGCTTGTGTTGGTGACGATCCTTGG
CAGGAGTTTTGTTGCATGGTTCGCAAAATCTTCATATACAGAAAGAGGAAGTGAGGAAG
ATGAACCCGGGGACTTTAAGCTGTAGGAGCGAGGAAGAAGCAGTTGTTGGGGAAGGATCA
GATGCAAGGACGCCAAGTCTGCATCAAATCCTTCATTGTCCAGCGCTGGGAACCTCTTAA
>G470 Amino Acid Sequence (domain in AA coordinates: 61-393)
MASSEVSMKGNRGGDNFSSSGFSDPKETRNVS VAGEGQKSNSR SAAERALDPEALYR
ELWHACAGPLVTVPRQDDRVFYFPQGHIEQVEASTNQAAEQMPLYDLPSKLLCRVINVD
LKAEDTDEVYAQITLLPEANQDENAIKEAPLPPPPRFQVHSFCKLTASDTSTHGGFS
VLRHRADECLPPLDMSRQPPTQELVAKDLHANERFRHIFRGQPRRHLLQSGWSVFVSSK
RLVAGDAFIFLNGENELRVGVRAMRQQGNVPSSVISSSHMLGLVATAWAHAISTGTMF
TVYYKPRTPSEFIVPFDQYMESVKNNYSIGMRFKMRFEGEAEPEQRFTGTIVGIEESDP
TRWPKSKWRSLKVRWDETSSIPRDRVSPWKVEPALAPPALSPVMPRPRKRPRSNIAPSS
PDSSMLTREGTTKANMDPLPASGLSRVLQGGQYESTLRTKHTESVECDAPENSVVWQSSAD
DDKVDVVS GSRRYGSENWMS SARHEPTYD LLSGFGTNIDPSHGQRIPFYDHSSSPSMPA
KRILSDSEGKFDYLANQWQMIHSGLSLKLHESPKVPAATDASLQGR CNVKYSEYPVLNGL
STENAGGNWPIRPRALNYEEVNAQAQAQAREQVTKQPFITQEEAKSREGNCR LFGIP
LTNNMNGTDSTMSQRNNLNDAGLTQIASPKVQDLSDQSKGSKSTNDHREQRPFQTN NP
HPKDAQTKTNSRSC TKVHKQGI ALGRSVDLSKFQNYEELVAELDRLFEFNGELMAPK D
WLI VYTDEENDMMLVGDDPWQEFCCMVRKIFIYTKEEVRKMNPGT LSCRSEEEAVVGE GS
DAKDAKSASNPSLSAGNS*
>G652 (1..606)
atgagcggaggagagacgtgaacatgagtggtggagacagacgcaagggaaacggtgaag
tggtttgatacacagaaggggttggtttcatcacacctagcgacggtggtgacgatctc
tcggttcaccagctctccatcagatctgaaggatttcgtagcctcgcagctgaggaatct
ggtgagttcgacgttgaggttgacaactccggccgtcccaaggctattgaagtgtctgga
cccagcgggtgctcccgttcagggttaacagcgggtggtggtggttcatctggtggacgcggg
gggttttgccggcggtggtggaagaggagggggacgtggtggaggaagctacggaggaggt
tatggtggaagaggaagcgggtggccgtggaggaggtggtggtgataattcttgctttaag
tgcggtgaaccaggtcacatggcgagagaatgctctcaagggtggtggaggatacagcggg
ggcggggggtggtggaaggtacgggtctggcgggcgcgaggaggaggtggtggtggctta
agctgctacagctgtggagagtctgggcactttgcaagggttgactagcgggtggtgct
cgttga
>G652 Amino Acid Sequence (domain in AA coordinates:28-49, 137-151, 182-196)
MSGGGDVNMSGDDRRKGT VKWFD TQKGFGFITPSDGGDDL FVHQSSIRSEGFRSLAAEES
VEFDVEVDNSGRPKAIEVSGPDGAPVQGN SGGGSSGGRGFGGGGGRGGRGGSYGGG
YGGRGSGRGGGGGD NSCFKCGEPGHMARECSQGGGGYSGGGGGGRYSGGGGGGGGGGL
SCYSCGESGHFARDCTSGGAR*
>G671 (61..1119)
TTCACCTTGAGAAACCCCTTTGAACTCGATCAAGAAAGCTAAGTTTGAAGAATCAAGA

ATGGTGCGGACACCGTGTGCAAAGCCGAAGTGGGTAAAGAAAGGAGCTTGGACTCCC
GAGGAAGATCAGAAGCTTCTCTTACCTTAACCGCCACGGTGAAGGTGGATGGCGAAGT
CTCCCCGAAAAAGCTGGACTCAAGAGATGCGGCAAAAGCTGCAGACTGAGATGGGCCAAT
TATCTTAGACCTGACATCAAAAGAGGAGAGTTCAGTGAAGACGAAGAAGCTTCAATCATC
TCTCTTCACGCCCTTCACGGCAACAAATGGTCTGCTATAGCTCGTGGACTACCAGGAAGA
ACCGATAACGAGATCAAGAACTACTGGAACACTCATATCAAAAAACGTTTGATCAAGAAA
GGTATTGATCCAGTTACACACAAGGGCATAACCTCCGGTACCGACAAATCAGAAAACTC
CCGGAGAAACAAAATGTTAATCTGACAACTAGTGACCATGATCTTGATAATGACAAGGCG
AAGAAGAACAACAAGAATTTTGGATTATCATCGGCTAGTTTCTTGAACAAAGTAGCTAAT
AGGTTTCGGAAGAGAATCAATCAGAGTGTCTGTCTGAGATTATCGGAAGTGGAGGCCCA
CTTGCTTCTACTAGTCACTACTAATACTACAACCTACAAGTGTTCCTGTTGACTCTGAA
TCAGTTAAGTCAACGAGTTCCTTCCCTTCGCACCAACCTCGAATCTTCTCTGCCATGGGACC
GTTGCAACAACACCAAGTTTCATCGAAGTTTGACGTTGATGGTAACGTTAATCTGACGTGT
TCTTCGTCACGTTCTCTGATTCTCCGTTAACAATCCTCTAATGTACTGCGATAATTTTC
GTTGGTAATAACAACGTTGATGATGAGGATACTATCGGGTCTCCACATTTCTGAATGAT
GAAGATTTTCATGATGTTGGAGGAGTCTTGTGTTGAAAACACTGCGTTCATGAAAGAACTT
ACGAGGTTTCTTCACGAGGATGAAAACGACGTCGTTGATGTGACGCCGGTCTATGAACGT
CAAGACTTGTTTGACGAAATTGATACTATTTTGATGAGTGAAGTCAATATCGATGAA
TCCCACGTGACCATGTCAATATGATGTCTATGGATATGTTACCTTGATGATGTTGATGGT
AATAATAATAAATAATAGATGGTGTATGATGACCATGCATGAATCATGAATGTAGTTCGTG
TTGTCACATATGCTTGTGTTTTTGTGTTTTTTTTTTTGGTCTGAAGTGTGTTGTTTCGT
TGTAATGATTATAAATGGTGTATGTAATAATTATAATGTTAAAAA

AAAA
>G671 Amino Acid Sequence (domain in AA coordinates: 15-115)
MVRTPCCKAELGLKKGAWTPEEDQKLLSYLNRHGEQWRTLPKAGLKRCKGKSCRLRWAN
YLRPDIKRGEFTEDEERSIISLHALHGNKWSAIARGLPGRDNEIKNYWNTHIKRLIKK
GIDPVTHKGITSGTDKSENLEKQNVNLTSDHDLNDKAKKNNKNFGLSSASFLNKVAN
RFGKRINQSVLSEIIGSGGGLASTSHTTNTTTSVSVDSSESVKSTSSSFAPTSNLLCHGT
VATTPVSSNFDVDGNVNLTCSSSTFSDSSVNPLMYCDNFVGNVNDDEDITIGFSTFLND
EDFMMLEESCVENTAFMKELTRFLHEDENDVDVTPVYERQDLFDEIDNYFG*

>G779 (110..712)
GACATGCATGTAAGCAATTCGGTTAATTAATCGAGTCAAAGATATATATCAGTAAATACAT
ATGTGTATATTTCTGAAAAAGAATATATATATTGAGAAATAAGAAAAGATGAAAATGGA
AAATGGTATGTATAAAAAGAAAGGAGTGTGCGACTCTTGTGTCTCGTCCAAAAGCAGATC
CAACCACAGCCCCAAAAGAAGCATGATGGAGCCTCAGCCTCACCATCTCCTCATGGATTG
GAACAAAGCTAATGATCTTCTCACACAAGAACACGCAGCTTTTCTCAATGATCCTACCA
TCTCATGTTAGATCCACCTCCCGAAACCTTAATTCACCTGGACGAAGACGAAGAGTACGA
TGAAGACATGGATGCGATGAAGGAGATGCAGTACATGATCGCCGTCATGCAGCCCGTAGA
CATCGACCTGCCACGGTCCCTAAGCCGAACCGCGTAACGTAAGGATAAGCGACGATCC
TCAGACGGTGGTTGCTCGTCCGGTCCGGAAAGGATCAGCGAGAAGATCCGAATTCCTAA
GAGGATCGTGCTGGTGGTGCGAAGATGGACACAGCTTCCATGCTCGACGAAGCCATACG
TTACACCAAGTTCTTGAAACGGCAGGTGAGGATTCTTCAGCCTCACTCTCAGATTGGAGC
TCCTATGGCTAACCCCTCTTACCTTTGTTATTACCACAACCTCCCAACCCGTATGAACCTAC
ACAGAAGCTCGCTAGCTAGACATTTGGTGTCTCCTCTCAACCTTT

>G779 Amino Acid Sequence (domain in AA coordinates: 126-182)
MKMENGMKKKGVCDSCVSSKSRSNHSPKRSMEPQPHLLMDWNKANDLLTQEHAAFLN
DPHHLMLDPPPETLIHLDEDEYDEMDAMKEMQYMIAMQPVDPATVPKPNRRNVRI
SDDPQTVVARRRRRERISEKIRILKRIVPGAKMDTASMLDEAIRYTKFLKRQVRILQPHS
QIGAPMANPSYLCYVHNSQP*

>G962 (148..1392)
CGTCGACTCTCTACTCAACCACTCAATTTTCATCTCTCTTTTCCCTTCCATTGTTAGT
ATAAAAACCAAGCAACCCCTTAATCACTTTTCATCATCATATATCACCTTAATCCACATG
CATACACATATCTAGTCTTTTGTATATATGGCAATGTATCCTCCACAACAAGCATCATT
CCCATGAGTAACCAAGTCAACAATAACGAAAAAGGTATAGAAGACAATGATCATAGAGGC
GGCCAAGAGAGTCATGTCCAAAATGAAGATGAAGCTGATGATCATGATCATGACATGGTC
ATGCCCCGATTTAGATTCCATCTACCGAAGAAGAACTCATAGAGTTTACCTTCGCCGA
AAAGTTGAAGGCAACGCTTTAATGTAGAACTCATCACTTTCCTCGATCTTTATCGCTAT

GATCCTTGGGAACCTTCCTGCTATGGCGGCGATAGGAGAGAAAGAGTGGTACTTCTATGTG
CCAAGAGATCGGAAATATAGAAATGGAGATAGACCGAACCGAGTAACGACTTCAGGATAT
TGGAAAGCCACCGGAGCTGATAGGATGATCAGATCGGAGACTTCTCGGCCTATCGGATTA
AAGAAAACCCTAGTTTTCTACTCTGGTAAAGCCCCCTAAAGGCACTCGTACTAGTTGGATC
ATGAACGAGTATCGTCTTCCGCACCATGAAACCGAGAAAGTACCAAAAGGCTGAAATATCA
TTGTGCCGAGTGTAACAAAGGCCAGGAGTAGAAGATCATCCATCGGTACCACGTTCTCTC
TCCACAAGACATCATAACCATAACTCATCGACATCATCCCGTTTAGCCTTAAGACAACAA
CAACACCATTATCCTCCTCTAATCATTCCGACAACAACCTTAACAACAACAACAACATC
AACAACTCTCGAGAAGCTCTCCACCGAATATTCCGGCGACGGCAGCACAACAACAACGACC
ACAAACAGTAACCTTGACGTTACCATTGCTCTAGCCAATCAAAACATATATCGTCCAATG
CCTTACGACACAAGCAACAACACATTGATAGTCTCTACGAGAAATCATCAAGACGATGAT
GAAACTGCCATTGTTGACGATCTTCAAAGACTAGTTAACTACCAATATCAGATGGAGGT
AACATCAATCACCAATACTTTCAAATTGCTCAACAGTTTCATCATACTCAACAACAAAAT
GCTAACGCAAAACGCATTACAATTGGTGGCTGCGGCGACTACAGCGACAACGCTAATGCCT
CAAACCTCAAGCGCGCTTAGCTATGAACATGATTCTGCAGGAACGATTCCAAACAATGCT
TTGTGGGATATGTGGAATCCAATAGTACCAGATGGAAACAGAGATCACTATACTAATATT
CCTTTTAAGTAATTTAATTAGATCATGATTATTATCCATGACAATAATTAATGCTGCTTT
GCGC

>G962 Amino Acid Sequence (domain in AA coordinates: 53-175)
MAIVSSTSIIPMSNQVNNEKGIENDHRGGQESHVQNEDEADDHDMVMPGFRFHPT
EEELIEFYLRKVEGKRFNVELITFLDLRYDPWELPAMAAIGEKEWYFYVPRDRKYRNG
DRPNRVTTSGYWKATGADRMIRSETSRPIGLKTLVFYSGKAPKGRTRTSWIMNEYRLPHH
ETEKYQKAEISLCRVYKRPVEDHPSVPRSLSTRHHNHSSTSSRLALRQQQHSSSSNH
SDNNLNNNNNINNLEKLSTEYSGDGSTTTTTTNSNSDVTIALANQNIYRPMYDTSNNTL
IVSTRNHQDDDETAIVDDLQRLVNYQISDGGNINHQYFQIAQQFHHTQQQNANANALQLV
AAATTATTLMPQTQAALAMNMIPAGTIPNNALWDMWNPIVDPGNRDHYTNIPFK*

>G977 (46..591)
CACCAAACCTACCTGAAACCCCTATTTCCATTTACCATTACACTAATGGCACGACCACAA
CAACGCTTTTCGAGGCGTTAGACAGAGGCATTGGGGCTCTTGGGTCTCCGAAATTTCGTCAC
CCTCTCTTGAAAACAAGAATCTGGCTAGGGACGTTTGAGACAGCGGAGGATGCAGCAAGG
GCCTACGACGAGGCGGCTAGGCTAATGTGTGGCCGAGAGCTCGTACTAATTTCCCATAC
AACCCCTAATGCCATTCTTACTTCTCTTCCAAGCTTCTATCAGCAACTCTTACCGCTAAA
CTCCACAAATGCTACATGGCTTCTCTTCAAATGACCAAGCAACGCAACACAAACGCAA
ACGCAGACCGCAAGATCACAATCCGCGGACAGTGACGGTGTGACGGCTAACGAAAGTCAT
TTGAACAGAGGAGTAACGGGAGACGACAGAGATCAAGTGGGAAGATGGAAATGCCAATATG
CAACAGAATTTTAGGCCATTGGAGGAAGATCATATCGAGCAAAATGATTGAGGAGCTGCTT
CACTACGGTTCCATTGAGCTTTGCTCTGTTTACCAACTCAGACGCTGTGAGAAATGGCC
TTGTCTGTTTTAGCGTATTCTTTTCATTTTTATTTTTGTTTCCACAAAAACGGCGTCGTA
GTGATGAGAGTAGTAGTGAGAGAAGGCTAATTTCAAGACATTTTGATCTGAATTGGCCTC
TTTTGAAACACTGATTCTAGTTTTCTATAAGAGCAATCGATCATATGCTATGTTATGTATA
GTATTATAAAAAAATGTTATTTTCTGATTNAAAAAAAAAAAAAAAAAAAAAAAAA

>G977 Amino Acid Sequence (domain in AA coordinates: 5-72)
MARPQQRFRGVRQRHWGSWVSEIRHPLKTRIWLTGFETAEDAARAYDEAARLMCGPRAR
TNFPYNPNAIPTSSSKLLSATLTAKLHKCYMASLQMTKQTQTQTQTARSQSADSDGVT
ANESHLNRGVTETTEIKWEDGNANMQNFRPLEEDHIEQMIEELLHYGSIELCSVLPTQT
L*

>G1063 (241..966)
GTTAAAGAAGATGGATGGGCCACAAGTTGCTATATAAATCCTTCCACTTCTTGTGTATA
CTATTGCTTGAGTTCTGATTGGGCACAGTAGTACCATTGCCATTTCTCTCACACATACCG
TCTCTTTCTCTCATCATCAATCATCAATCATCCAAAAGAAAAACCCCTAAAATTTCACTT
GTAAGCTTTTACCAGTTTCTCTCCATACCCATTTTATCAGCTTCTCCATATCTTTCTCT
ATGGATTCTGACATAATGAACATGATGATGCATCAGATGGAGAAGCTTCTGAGTTTGT
AACCCCTAATCCTCTTTCTCTCTCCGACCACAACAACACTTACCCTTTTCTCTTTAAC
TCCACTCATTACCAGTCCGATCACTCAATGACCAACGAACAGGTTTCCGCTACGGTTCC
GGTTTACTCACTAACCCCTTCTCTATCTCTCCCAACACAGCTTACTCTTCCGTTTTTCTT
GACAAAAGAAACAACAGTAACAACAACAATAATGGCACGAACATGGCAGCTATGCGAGAG
ATGATCTTCCGTATCGCCGTGATGAACCGATCCATATCGATCCCGAGGCGGTTAAGCCA

CCGAAGAGGAGGAACGTCAGGATCTCTAAAGATCCTCAAAGCGTGGCGGC'TAGGCATAGA
AGGGAGAGAATAAGCGAGAGGATTTCGGATTTTGAACGGCTTGTTCCTGGTGGGACGAAG
ATGGATACAGCTTCGATGCTCGATGAAGCAATTCATTATGTGAAGTTTTTAAAGAAACAG
GTGCAGTCTCTGGAGGAGCAGGCGGTGGTTACTGGCGGAGGGGAGGAGGAGGAGGAAGG
GTTTTGATCGGTGGAGGTGGAATGACGGCGGCGAGTGGTGGTGGTGGCGGGGGAGTG
GTTATGAAAGGGTGTGGAACAGTGGGGACTCATCAGATGGTGGGCAATGCACAGATTCTT
AGATGATGATGATTTTAAATTTTATTATTATTAATGTTGGAGAAAAAGAGAAAAA
TGATTCTGGAGAGGAAGCCAAGTAATTTATGTGAGAGTCTTTAATTTAACTTTATTTTC
TTGTTTAGATAATGTGTAATGATGGTTTTTAAAGCCAAAGACTCTCCATGGTTGTGGAG
CGAGTTTG

>G1063 Amino Acid Sequence (domain in aa coordinates: 131-182)

MDSDIMNMMMHQMEKLPEFCNPNSFFSPDHNNTPFLFNSTHYQSDHSMTNEPGFRYGS
GLLTNPSSISPNTAYSSVFLDKRNNNNNNNGTNMAAMREMIFRIAVMQPIHIDPEAVKP
PKRRNVRIKDPQSVAAHRHRRERISERIRILQRLVPGGTKMDTASMLDEAIHYVKFLKKQ
VQSLLEEQAVVTGGGGGGGGRVLIGGGGMTAASGGGGGGGVVMKCGTVGTHQMVGNAQIL
R*

>G1140 (67..729)

ATCCAAGATCCTCCAACCTCACAGAAAGGCAGATTCAAGAACAGTAGTGAAGGAGAGATCT
GGTAAATGGCGAGAGAGAAGATAAGGATAAAGAAGATTGATAACATAACAGCGAGACAA
GTTACTTTCTCAAAGAGAAGAAGAGGAATCTTCAAGAAAGCCGATGAACCTTTTCAGTTCTT
TGCGATGCTGATGTTGCTCTCATCATCTTCTCTGCCACCGGAAAGCTCTTCGAGTTCTCC
AGCTCAAGAATGAGAGACATATTGGGAAGGTATAGTCTTCATGCAAGTAACATCAACAAA
TTGATGGATCCACCTTCTACTCATCTCCGGCTTGAGAATTGTAACCTCTCCAGACTAAGT
AAGGAAGTCGAAGACAAAACCAAGCAGCTACGGAAGTGAAGAGGAGAGGATCTTGATGGA
TTGAACTTAGAAGAGTTGCGAGCGCTGGAGAACTACTTGAATCCGGACTTAGCCGTGTG
TCTGAAAAGAAGGGCGAGTGTGTGATGAGCCAAATTTTCTCACTTGAGAAACGGGGATCG
GAATTGGTGGATGAGAATAAGAGACTGAGGGATAAACTAGAGACGTTGGAAGGGCAAAA
CTGACGACGCTTAAAGAGGCTTTGGAGACAGAGTCCGGTGACCACAAATGTGTCAAGCTAC
GACAGTGAAGTCCCTTTGAGGATGACTCCGACACTTCCCTGAAGCTTGGGCTTCCATCT
TGGGAATGAATCTGAGAGAGAGAAAGATCCAGCAGAGTTGACTTCGATGGAAGCCCACAA
ATATTAAGTCTACCTTTTCCCTTTCTTTTCTTTGAATAAGTGTTGAAAAAGAATTGAGAT
GGGAAGGATGAATTCTCATTGCATTGCAGAGAAGCAAGTTTCAGATATTGTACGTGTTAT
TGGGTCTTTATAACTATTTTCTCCCCAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G1140 Amino Acid Sequence (conserved domain in AA coordinates:2-57)

MAREKIRIKKIDNITARQVTFKRRRGIFKKADELVLCDADVALIIFSATGKLFEFSSS
RMRDILGRYSLHASNINKLMDPPSTHLRLNLCNLSRLSKEVEDKTKQLRKLRGEDLDGLN
LEELQRLKLLSGLSRVSEKKGECVMSQIFSLEKRGSELVDENKRLRDKLETALERAKLT
TLKEALETESVTTNVSSYDSGTPLEDDSDTSLKLGLPSWE*

>G1425 (43..1005)

ACTCTCTCAAACCATAAAAAATATTCTCCGATCATCATTTTAAATGGAGAGTACAGATTCT
TCCGGTGGTCTCCGCGCGCGCAACCAACCTCCCTCCAGGATTCCGGTTTCATCCAACA
GACGAAGAACTTGTAATTCATTACCTCAAACGCAAAGCAGATTCTGTTCTTTTACCAGTC
GCGATCATCGCCGACGTTGATCTTTACAAATTTGATCCATGGGAAC'TCCCGCGAAAGCT
TCGTTTGGGAGAACAAGAATGGTATTTTTTTCAGTCCAAGAGATCGGAAATATCCCAACGGA
GCTAGACCTAACCGAGCTGCGACTTCCGGTTATTGGAAAGCGACTGGTACAGATAAACCG
GTGATTTCAACCGCGCGGTGGTGGTAGTAAAAAAGTGGGAGTTAAAAAGGCTCTAGTGTTT
TACAGTGGTAAACCACCAAAAGGAGTTAAATCAGATTGGATTATGCATGAATATCGGTTA
ACTGATAATAAACCTACTCACATTTGTGACTTCGGCAACAAGAAAACTCTCTCAGGCTT
GATGATTGGGTGTTGTGTCGTATCTACAAGAAAAACAATAGTACAGCATCTAGACATCAT
CATCATCTTCATCATATTATCATCTAGATAATGATCATCATCGTCATGATATGATGATTGAT
GATGATCGATTCCGTCATGTTCTCTCGGTCTTCACTTCCCGGCGATTTTTTCTGACAAT
AATGATCCGACGGCTATATATGATGGTGGCGGCGCGGATACGGAGGTGGAAGTTACTCG
ATGAATCATTTGTTTCGCATCTGGATCAAAGCAGGAGCAGTTGTTTCCACCGGTGATGATG
ATGACTAGTCTAAATCAAGATTCCGGTATTGGATCGTCGTCTCGTCACCTAGCAAGAGATTT
AACGGCGGCGCGGCTTGGAGATTGTTCCGACTTCTATGGCGGCGACGCCGTTAATGCAGAAC
CAAGGTGGGATTTACCAATTGCCTGGTTTGAATTGGTATTCTTGAAAACAATTTACGATG
AAGAATTTTTAAATTTGTGTATATATACGGTTTGAGTGATTAGGGGGCATTGGGGGA

TTTATTTACGGTTGATTATTATTGTAGTGTATAGAACTAAGGAGATTAAATTAAATAGA
TTGGAGGAAAAAAAAAAAAAAAAA

>G1425 Amino Acid Sequence (domain in AA coordinates: 20-173)
MESTDSSGGPPPPQPNLPPGFRHPTDEELVIHYLKRKADSVPLPVAIADVDLYKFDPW
ELPAKASFGEQEWYFFSPDRKYPNGARPNRAATSGYWKATGTDKPVISTGGGSKKVG
KKALVFYSGKPPKGVKSDWIMHEYRLTDNKPHTICDFGNKKNSLRLLDDWVLCRIYKKNS
TASRHHHLHHIHLNDNDHHRHDMIDDDRFRHVPPGLHFPALFSDNNDPTAIYDGGGGGY
GGGSYSNMNHCFASSGSKQEQLFPPVMMMTSLNQDSGIGSSSSPSKRFNGGGVGD CSTSMAA
TPLMQNQGGIYQLPGLNWYS*

>G1449 (105..581)
TAGACAGAGAGAAATAGAGAGAGAGACATGAAGAGCACTCTCAATAGAGAAG
AGAAGGAAGCATGAAGCTAGCTCTGCAGCTTCAAGGTCTCATTAAATGGAGGTCTCTAACT
CTTGTTCTTCATTTTCTTCATCCTCTGTCGACAGTACTAAACCTTCTCCTTCTGAATCTT
CTGTTAATCTCTCCCTTAGTCTCACATTTCTTCTACTTCTCCACAAAGAGAAGCAAGAC
AAGATTGGCCACCGATAAAGTCTAGATTAAGAGATACACTAAAGGGTCGTCGTCTTCTTC
GTCGTGGTGATGACACTTCTCTCTTTGTTAAGGTTTATATGGAAGGTGTTCCCATTTGGAA
GAAAACTCGACCTTTGCGTATTCTCAGGCTACGAGAGTCTATTAGAAAATCTCTCTCACA
TGTTTCGATACTTCAATCATCTGCGGTAATCGAGATCGAAAACATCATGTTTTGACATATG
AAGACAAGGATGGAGATTGGATGATGGTCGGAGATATTCATGGGATATGTTTCTTGAAA
CCGTGAGAAGACTAAAGATCACGAGACCGGAGAGGTATTAAGGTTGGATCGGTCAAGGC
TGTGATTGCGCAGTTACGAGACGTGTAAGATTTAGGCATTGATGAAGAGACTTGAGGCGG
GACGGAGCTATTGCTGCATATTGCAACAAAGGCCTTGAAGAAGTTGGAGAATTGATTGAT
GCATATATTTATTTATATGACACCTTTGAGTGTGTTTTTCTTATAAATAAATCACAATA
TCCAAGACTTCTCTTTAAA

>G1449 Amino Acid Sequence (domain in AA coordinates: 48-53,74-107,122-152)
MEVSNSSCSFSSSSVDSTKPSPESSVNLSSLTFPSTSPQREARQDWPPPIKSRLRDLTK
GRRLLRGGDDTSFLVKVMEGVPIGRKLDLCVFSGYESLLENLSHMFDTSIICGNRDRKH
HVLTYEDKGDWMMVGDIPWDMFLETVRRLKITRPERY*

>G1897 (1..678)
ATGCCTTCTGAATTCAGTGAATCTCGTTCGGGTTCTTAAGATTCCCCACGGCCAAGGAGGA
TCTGTTGCGATTCCGACGGATCAACAAGAGCAGCTTTCTTGTCTCGCTGTGAATCAACC
AACACCAAGTTCTGTACTACAACAACCTACAACCTCTCACAACCTCGTCATTTCTGCAAG
TCTTGTGCGCGTTACTGGACTCATGGAGGTACTCTCCGTGACATTCCTCGTGGTGTGTT
TCCCGTAAAAGCTCAAAACGTTCCCGGACTTATTCCTCTGCCGCTACCACCTCCGTTGTC
GGAAGCCGGAACCTTCCCTTACAAGCTACGCTGTCTTTTCCCTCAGTCGTCTTCCAAC
GGCGGTATCACGACGGCGAAGGGAAGTGCTTCGTCTGTTCTATGGCGGTTTCAGCTCTTTG
ATCAACTACAACGCGCGGTGAGCAGAAATGGGCTGGTGGCGGGTTAATGGGCCAGAT
GCTTTTGGTCTTGGGCTTGGTTCAGGGTCTGTTATGAGGACGTCAGATATGGGCAAGGA
ATAACGGTCTGGCGGTTTCAAGTGGCGCTACTGATGCTGCAACTACTACAAGCCACATT
GCTCAAATACCGCCACGTGGCAGTTTGAAGGTCAAGAGAGCAAAGTCGGGTTCTGTCT
GGAGACTACGTAGCGTGA

>G1897 Amino Acid Sequence (domain in AA coordinates:34-62)
MPSEFSESRRVPKIPHGQGSVAIPTDQEQQLSCPRCESTNTKFCYNNYNFSQPRHFCK
SCRRYWTHGGTLRDPVGGVSRKSSKRSRTYSSAATTSVVGSRNFPLOATPVLFPQSSN
GGITTAKGSASSFYGGFSSLINYNAAVSRNGPGGGFNGPDAGLGLGHGSYYEDVRYGQG
ITVWPFSSGATDAATTSIAQIPATWQFEGQESKVG FVSGDYVA*

>G2143 (89..784)
TCTTCTTCTCTCTCCATACCTTATCTCACCAGCTTCTCCATATCTCTCAAAGAAAAACA
AACCTTATAAATCCACAAAAAGGAGGATGGATAACTCCGACATTCTAATGAACATGAT
GATGCAGCAGATGGAGAAGCTTCTGAACACTTCTTAACCTCAAACCTAACCCTAATCC
CCATAACATTATGATGCTTTCTGAATCCAACCCACCCGTTCTTCTTCAACCCCACTCA
TTCTCATCTCCCATTTGACCAAACCATGCCTCACCACCAACCCGTTTAAATTTCCGGTA
CGCCCCCTCCCGTCATCATCTCTCCCGGAGAAGAGAGGAGGCTGCAGCGACAACGCCAA
CATGGCGGCGATGAGAGAGATGATCTTTCGAATAGCCGTGATGCAGCCTATACATATTGA
TCCCGAATCCGTAAAGCCACCAAAGAGAAAGAGAGGATCTCTAAGGATCCACAGAG
CGTGGCAGCTCGGCATCGAAGGGAGAGGATAAGCGAGCGGATTCCGATTCTTCAGCGGCT
TGTTCCCGGTGGGACTAAGATGGATACGGCGTCGATGCTCGATGAGGCTATCCATTACGT

TAAGTTTCTCAAGAAGCAAGTGCAGTCGCTGGAGGAACATGCGGTGGTTAACGGCGGAGG
 AATGACGGCGGTGGCCGGAGGAGCACTTGCGGGTACTGTTGGTGGAGGATATGGAGGAAA
 AGGGTGTGGCATTATGCGGTCTGATCATCACCAGATGCTTGGAAATGCACAGATTCTTAG
 ATGATGATGATGTTGATTTTTAAATATATATCATATGTTTATTAATATGACGGGAAAAAA
 TATTATCGAGGGAGTTGAATTTAGTATCATGAAACTATGAGAGCATTTTTTTAAATGTT
 TTTATCTTTCCGGGTTTCGATAATGTTTGGGATGGTTAATTAACAATTTAAAAGTCAGAC
 AACTTGGTTGTAAAGACTAAAGAATAAGCATAGTTTATCAATTTATCATTACTAAATGAA
 ATAG

>G2143 Amino Acid Sequence (domain in aa coordinates: 128-179)

MDNSDILMNMQQMEKLPFHFSNSNPNNPHNIMMLESNTHPFFNPTHSHLPFDQTM
 PHHQPLNFRYAPSPSSSLPEKRGGCSDNANMAAMREMI FRIAMQPIHIDPESVKPPKR
 KNVRISKDPQSVAAHRHRERISERIRILQRLVPGGKMDTASMLDEAIHYVKFLKKQVQS
 LEEHAVVNGGGMTAVAGGALAGTVGGGYGGKGGIMRSDHHQMLGNAQILR*

>G2535 (1..1005)

ATGAACATATCAGTAAACGGACAGTCACAAGTACCTCCTGGCTTTAGGTTTCACCCAACC
 GAGGAAGAGCTCTTGAAGTATTACCTCCGCAAGAAAATCTTAACATCAAGATCGATCTC
 GATGTTATTCCTGACATTGATCTCAACAAGCTCGAGCCTTGGGATATTCAAGAGATGTGT
 AAGATTGGAACGACGCCGCAAAACGATTGGTACTTTTTATAGCCATAAGGACAAGAAGTAT
 CCCACCGGGACTAGAACCAACAGAGCCACCACGGTCGGATTTTGGAAAGCGACGGGACGT
 GACAAGACCATATATACCAATGGTGATAGAAATCGGGATCGGAAAGACGCTTGTCTTCTAC
 AAAGGTCGAGCCCCCTCATGGTCAGAAATCCGATTGGATCATGCACGAATATAGACTCGAC
 GAGAGTGATTAATCTCCTCGTGTGGCGATCATGACGTCAACGTAGAAACGTGTGATGTC
 ATAGGAAGTCGCAAGGATGGGTGGTGTGTCGTGTTTTCAAGAAAAATACCTTTGCAAA
 AACATGATTAGTAGTAGCCCGCGAGTTCGGTGAAAACGCCGTCGTTCAATGAGGAGACT
 ATCGAGCAACTTCTCGAAGTTATGGGGCAATCTTGTAAGGAGAGATAGTTTTAGACCCT
 TTCTTAAAACTCCCTAACCTCGAATGCCATAACAACACCACCATCACGAGTTATCAGTGG
 TTAATCGACGACCAAGTCAACAACCTGCCACGTCAGCAAAGTTATGGATCCCAGCTTCATC
 ACTAGCTGGGCCGCTTTGGATCGGCTCGTTGCCTCACAGTTAAATGGGCCCCAATCGTAT
 TCAATACCAGCCGTTAATGAGACTTCACAATCACCGTATCATGGACTGAACCGGTCCGGT
 TGTAAATACCGGTTTAAACACAGATTACTATATACCGGAGATTGATTTATGGAACGAGGCA
 GATTTTCGCGAGAACGACATGCCACTTGTGTAACGGTAGTGGATAA

>G2535 Amino Acid Sequence (conserved domain in AA coordinates: 11-114)

MNISVNGQSQVPPGFRFHPTEBELLKYLYLRKKISNIKIDLDVIPDIDLNKLEPWDIQEMC
 KIGTTPQNDWYFYSHKDKICYPTGTRTNRATTVGFWKATGRDKTIYINGDRIGMRKTLVFI
 KGRAPHGQKSDWIMHEYRLDESVLISSCGDHVDNVETCDVIGSDEGWVVCRVFKNNLCK
 NMISSPASSVKTPSFNEETIEQLLEVMGQSCKGEIVLDPFLKLPNLECHNNTTITSYQW
 LIDDQVNNCHVSKVMDPSFITSWAALDRLVASQLNPGNSYSIPAVNETSQSPYHGLNRS
 CNTGLTPDYIPEIDLWNEADFARTTCHLLNGSG*

>G2557 (94..1215)

TCGACTTCCTGTGAACATCTGTTTGTCTCTTCTTCCGGTTTCACTTTTTTCATGTCCT
 GCCGTTATTACAACGAGGATTGTGTTTGATCCGATGGAAGGATTGGAATCTGTGTACGCT
 CAAGCTATGTATGGAATGACACGAGAGAGCAAAATCATGGAGCATCAAGGATCAGATTTG
 ATTTGGGGAGGAAATGAGCTAATGGCTCGAGAACTCTGTTCTTCTTCTTCTTATCACCAC
 CAACTCATTAATCCGAATCTTAGCAGCTGTTTCATGTCTGATCTTGGAGTCTTAGGTGAG
 ATTCAACAGCAGCAACATGTTGGCAACAGAGCTAGCTCGATAGATCCATCATCACTCGAT
 TGTTTGTATCTGCGACGTGCAATAGCAACAACACCTCGACGGAGGACGATGAAGGAATA
 TCTGTGCTTTTCTCAGATTGTCAGACTCTTTGGAGCTTTGGTGGAGTCTCATCTGCAGAG
 TCTGAGAACAGAGAGATCACTACTGAGACGACAACAACGATAAAGCCTAAGCCTTTGAAG
 AGAAACAGAGGAGGAGATGGAGGAACTACTGAGACTACAACAACAACAAAACCTAAG
 TCTTTGAAGAGAAACAGAGGAGACGAGACAGGAAGTCACTTAGTCTTGTTCATCCTCAA
 GATGATTCGGAGAAAGGAGGTTTCAAGCTTATATACGATGAGAATCAATCGAAATCAAAG
 AAACCAAGAACAGAGAAAGAACGAGGCGGTTCTTCAACATTAGTTTCCAACATTCAACT
 TGTTTGTCTGACAATGTCGAGCCCGATGCTGAGGCGATTGCACAAATGAAGGAGATGATA
 TACAGAGCGGCTGCATTTAGACCGGTGAATTTCCGGTTAGAGATTGTGGAGAAGCCTAAG
 AGGAAGAACGTCAAGATATCGACGGATCCTCAAACGGTTGAGCGAGACAGAGAAGGGAG
 AGGATAAGTGAGAAGATTAGGGTTTTACAAACATTGGTTCCAGGTGGGACGAAGATGGAT
 ACTGCATCAATGCTTGATGAAGCTGCTAATTATCTCAAGTTTCCTTAGAGCACAAAGTAAAA

GCTTTTAGAAAACCTTGAGACCCAAGCTTGACCAAACCAATCTCTCTTTCTCTTCTGCTCCT
ACATCGTTTCCATTATTCCACCCATCTTTCTTCCATTGCAAAATCCTAATCAAATCCAT
CATCCAGAGTGTGACAGATTATAAACTTTTGAGTTTCATCATCATCAACAGAATCATGG
CGTCTTGATTGTTTTAGCAGTTCTCAAGAAAGGCAACTTCTGTGACAAGGGTGGTGTGCG
GCAGTGTGTTTACACTTTCCAGTCTTTGTTTTGCATTTCTTTTATATAAAAGTTTGTAT
TTTATATAGAATCTGTGGAATTCGAGGGTTGAAATATTGTGAAAAACAGAGCCGCAAGAG
GTTAATTACAGTCTCTGCAATATTTCAACCTTTTATTACTTTATTAGAGTAAAGATAGC
GT

>G2557 Amino Acid Sequence (domain in aa coordinates: 278-328)

MEGLSVYAQAMYGMTRESKIMEHQGSDLIWGGNELMARELCSSSSYHHQLINPNLSSCF
MSDLGVLGELIQQQHVGNRRASSIDPSSLDCLLSATNSNNSTEDDEGISVLFSQCQLW
SFGGVSSAESENREITTTETTTTIKPKPLKRNRRGGDGGTTETTTTTPKPSLKRNRRDET
SHFSLVHPQDDSEKGGFKLIYDENQSKSKPRTEKERGGSSNISFQHSCTCLSDNVEPDAE
AIAQMKEMIYRAAAFPRVNFGLIEVKPKRKVKISTDPQTVAARQRRERISEKIRVLQT
LVPGGTKMDTASMLDEAANYLKFLRAQVKALENLRPKLDQTNLSFSSAPTSFPLFHPSPFL
PLQNPNIHHPEC*

>G259 (52..786)

GAGATCTTCTACTACTTGTGTTTTCTCAAGAATAATAATTTTCGTTTATATATGGAAGAT
GCTGGTGAACATTTACGGTGTAAACGATAACGTTAACGACGAGGAGCGTTTGCCATTGGAG
TTTATGATCGGAACTCAACATCCACGGCGGAGCTACAGCCGCTCCACCGTTCTTGGA
AAGACATACAAAGTGGTGGAGGATCCGACGACGGAGGGTTATATCTTGAACGAATAC
GGAAGTGGTTTTCGTCGTGTGGCAGCCGGCAGAAATTCGCTAGAGATCTGTTACCAACTT
TTCAAGCATTGCAACTTCTCTAGCTTCGTTCCGACGCTCAATACTTACGGTTTTCGAAAA
GTAACGACGATAAGATGGGAATTTAGTAATGAGATGTTTCGAAAGGGGCAAAGAGAGCTT
ATGAGCAATATCCGAAGAAGGAAGCCAACATTTGGTCACACAACAAGTCTAATCACCAG
GTTGTACCAACAACAACGATGGTGAATCAAGAAGGTCATCAACGGATTGGGATTGATCAT
CACCATGAGGATCAACAGTCTTCCGCCACTTCATCCTCTTTTCGTATACACTGCATTACTC
GACGAAAAACAAATGCTTGAAGAATGAAAACGAGTTATTAAGCTGCGAACTTGGGAAAACC
AAGAAGAAATGCAAGCAGCTTATGGAGTTGGTGGAGAGATACAGAGGAGAAGACGAAGAT
GCAACTGATGAAAGTGATGATGAAGAAGATGAAGGGCTTAAGTTGTTCCGGAGTAAACTT
GAATGAAACTAGATTGCTAGATTGATATTCTGTAATATACCAGTTTCTTCATATTCTTAGA
AGTTTTGCATACTATATATAGTACTCTTTTAAGACATGCAAGATCAGAACATATG

>G259 Amino Acid Sequence (domain in AA coordinates: 27-131)

MEDAGEHLRCNDNVNDEERLPLEBFMIGNSTSTAELOPPPPFLVKYKVVEDPTTDGVISW
NEYGTGFVWQPAEFARDLLPTLFKHCNFSFVRQLNTYGRKVTITIRWEFSNEMFRKGQ
RELMSNIRRRKSQHWSHNKSNNHQVPTTTMVNQEGHQRIQIDHHHEDQSSATSSSFVYT
ALLDENKCLKNENELLSCELGKTKKKCKQLMELVERYRGEDEDATDESDDDEDEGLKLF
VKLE*

>G353 (82..570)

ACCAAACCTCAAAAAACACAAACCACAAGAGGATCATTTTCATTTTTATTGTTTCGTTT
ATCATCATCATCAGAAGAAAAATGGTTGCGATATCGGAGATCAAGTCGACGGTGGATGTC
ACGGCGGCGCAATTGTTTGATGCTTTTATCTAGAGTTGGACAAGAAAACGTTGACGGTGGC
GATCAAAAACGCGTTTTTACATGTAAAACGTGTTTGAAGCAGTTTCATTTCGTTCCAAGCC
TTAGGAGGTACCGTGCGAGTCACAAGAAGCCTAACAACGACGCTTTGTCGTCTGGATTG
ATGAAGAAGGTGAAAACGTCGTCGCATCCTTGTCCCATATGTGGAGTGGAGTTTCCGATG
GGACAAGCTTTGGGAGGACACATGAGGAGACACAGGAACGAGAGTGGGGCTGCTGGTGGC
GCGTTGGTTACACGCGCTTTGTTGCCGAGCCACGGTGACTACGTTGAAGAAATCTAGC
AGTGGGAAGAGAGTGGCTTGTGATCTGAGTCTAGGGATGGTGGACAATTTGAATCTC
AAGTTGGAGCTTGAAGAACAGTTTATTGATTTTATTTATTTTCCTTAAATTTTCTGAAT
ATATTTGTTTCTCTCATTCTTTGAATTTTCTTAAATATTCTAGATTATACATACATCCGC
AGATTTAGGAACTTTCATAGAGTGAATCTTTTCTTTCTGTAAAAATATATTTTACTTG
TAGCAAA

>G353 Amino Acid Sequence (domain in aa coordinates: 41-61, 84-104)

MVAISEIKSTVDVTAANCLMLLSRVGQENVDDGQKRVFTCKTCLKQPHSFQALGGHRAS
HKKPNNDALSSGLMKKVKTSHPICGVEFPMQALGHHMRRHRNESGAAGGALVTRAL
LPEPTVTTLKSSSGKRVACLDLSLGMVDNLNLKLELGRITVY*

>G354 (27..533)

CCTAGAAGTCACTAAGTCGATTCAAAATGGTTGCGAGAAGTGAGGAAATTGTGATAGTGG
AAGAAGATACGACTGCGAAATGTTTGTATGTTTATCAAGAGTCGGAGAATGCGGCGGCG
GCTGCGGGGAGATGAACGTGTTTTCCGATGCAAGACTTGTCTTAAAGAGTTCTCATCGT
TTCAAGCTTTGGGAGGTCATCGTGCAAGCCACAAGAACTTATCAACAGTGACAATCCAT
CACTTCTTGGATCCTTGTCCAACAAGAAAATAAAACGTCTCATCCTTGTCCGATATGTG
GAGTGAAGTTTCCGATGGGACAAGCTTGGTGGTCACATGAGGAGACATAGGAACGAGA
AAGTCTCAGGCTCGTTGGTTACACGTCTTTTCTACCGGAGACGACGACGGTGACGGCTT
TGAAGAAATTTAGTAGTGGGAAGAGAGTGGCTTGTGGATTGGACTTAGATTTCGATGG
AGAGTTTGGTCAATTGGAAGTTGGAGTTGGGAAGAACGATTCTTGGAGTTAAGTTTGTG
GGTTGTATACAGTTTACATGATTTTGTATCTTTGTGATCCAATTATCGTACCGATCG
ATGTGAATATTAATTTTGATACAATAAAA

>G354 Amino Acid Sequence (domain in aa coordinates: 42-62, 88-109)

MVARSEEIVIVEEDTTAKCLMLLSRVGECGGGCGDERVFRCKTCLKEFSSFQALGGHRA
SHKKLINSNDNPSLLGSLSNKKTSHPCPICGVKPFMGQALGGHMRRHRNEKVSGSLVTR
SFLPETTTVTALKKFSSGKRVACLDDLDSMESLVNWKLELGRITISWS*

>G638 (86..1861)

GAATTAAGGTTTAACTTTTACCTTTTTTCCCTTCACTATCGATAATTGATCTTCTCT
TTCGGCTGAATATAAATCTGAAAAAATGGATCAAGATCAGCATCCTCAGTACGGTATACC
GGAGCTCCGGCAGCTCATGAAAGGCGGAGGAAGGACGACTACTACAACACCGTCTACTTC
TTCTCATTTTCCCTCTGATTCTTTCGGTTTAACTTGTCTCCGGTGCAGCCACCGCCACA
CCGCTTTCATCAGTTCACTACTGATCAAGATATGGGTTTCTTGGCACGTGGCATAACATGG
ATTGGGTGGAGGTTCTTCAACGGCTGGAATAACAGTAACCTTAAACCGCAGTACTAGTGG
TGGAGGAGTTGGGTTTAGTGGGTTTCTTGACGGTGGTGGTTTCGGCAGCGGAGTAGGAGG
AGACGGTGGAGGAAGTGAAGGTGGCCGAGACAAGAAACCCTAACTCTGTTGGAAATTAG
ATCTCGTCTTGTATCATAAATTCAAAGAAGCTAATCATAAAGGACCTCTTTGGGATGAAGT
TTCTAGGATTATGTCCGAGGAACATGGATACCAAAGGAGTGGGAAGAAATGCAGAGAGAA
GTTTGAGAATCTGTACAAATACTATAGTAAGACTAAAGAAGCGGAAGCCGGAAGACAAGA
CGGAAACATCACAGATTTTTCCGCCAGCTCCAAGCGCTATACGGGGATTCTAATAACTT
GGTTTCTTGTCCCAATCATAACACGCGATTCATGAGCAGTGCTCTTCATGGTTTCCATAC
TCAAACCCATGAACGTTGCTACAACAACGTCACATCCATAACGTTGATAGTGTTC
TGGTTTTTCATCAAAGCCTTAGTCTTCTAACAACCTACAACCTCCTCCGAGCTTGAGCTGAT
GACTTCTCTTTCGGAAGGGAATGATTCTAGTAGTAGAAGGAAAAAGAGGAGTTGGAAGC
GAAGATAAAGGAGTTTATTGATACGAACATGAAAGGTTGATAGAGAGGCAAGATGTTTG
GCTTGAGAAGTTGACAAAGGTTATTGAAGACAAAGAGGAACACGGATGATGAAGAAGA
GGAATGGAGGAAGATTGAAGCTGCAAGGATTGATAAAGAGCATTTGTTTTGGGCTAAAGA
GAGGGCGAGGATGGAAGCTAGGGATGTTGCGGTGATTGAGGCATTGCAATACTTGACAGG
AAAGCCATTGATAAAGCCGCTGTGTTTATCCCGGAAGAGAGGACAAATGGTAATAATGA
GATCCGAAACAATAGTGAGACACAGATGAGAATGGAAGCGATCAAACGATGACTAACAA
TGTTTGTGTTAAAGGAAGTAGTAGCTGCTGGGGTGAGCAAGAGATTTTAAAGCTTATGGA
GATAAGAACCAGCATGGACTCGACCTTTCAAGAGATATTAGGAGGGTGCTCGGATGAGTT
TCTATGGGAGGAAATCGCAGCGAAGTTGATTCAAGTTAGGGTTTATCAGAGAGAGTGCCTT
ATTATGCAAGGAAAAGTGGGAATGGATAAGCAATGGAATGAGGAAAGAAAAGAACAAAT
CAACAAGAAAAGAAAGGATAATTTCGTCCAGCTGCGGCGTGTACTACCCGAGAAACGAAGA
AAATCCAATCTACAATAATCGAGAAAGTGGATATAATGATAATGATCCGCATCAAATCAA
CGAACAAGGCAATGTAGGTTCTTCAACATCAAACGCAAACGCAAACGCAAACGTAACCAC
TGGAAATCCGAGCGGTGCAATGGCTGCTAGTACAACTGCTTCCCGTTCTTCATGGGAGA
TGGAGATCAGAATTTGTGGGAGAGTTATGGTTTGAGGCTCAGTAAAGAAGAGAATCAGTA
AGTAATTTCTCTTAATGAAGAAGAAGAAGGTAATCATGTGGTTAACTAATTCTTTTGTG
TAGCTATATATGAGATAAACCTTGACTTAGCTATTATATGTCACATGCTGCTTAGAATTA
AGAAATATTTGTTGGGGCTTAACGAATTATATATCAGCATATATAAGATGAGAGTCTAAG
AATTATATCAAATTAGGCTTTAACCAACGTACGATTATATATTATGTTTTCATGTATTTA
TTCTGTAAGACTTTTTAATATCAATCTTTCTCTAAA

>G638 Amino Acid Sequence (domain in AA coordinates: 119-206)

MDQDQHPQYGIPELRQLMKGGGRTTTTTPTSSHPSPDFGFNLAPVQPPPHRLHQFTTD
QDMGFLPRGIHGLGGSSSTAGNNSNLNASTSGGGVGFSGFLDGGGFGSGVGGDGGGTGRW
PRQETLTLLEIRSLDHKFKEANHKGPLWDEVSRIMSEEHGYQSRGKKCREKFENLYKYY
SKTKEGEAGRQDGKHHFRQLQALYGDSSNNLVSCPNHNTQFMSSALHGFHTQNPMPNVAT

TTSNIHNVDSVHGFHQSLSLSNYNSSSELELMTSSSEGNDSRRRKRKRSWKAKIKEFIDT
NMKRLIERQDVWLEKLTQVIEDKEEQRMKEEWRKIEAARIDKEHLFWAKERARMEARD
VAVIEALQYLTKGKPLIKPLCSSPEERTNGNNEIRNNSETQNGSDQTMNNVCVKGSSS
CWGEQEILKLMEIRTSMDSTFQEILGGCSDEFLWEEIAAKLIQLGFDQRSALLCKEKWEW
ISNGMRKEKKQINKKRKDNSSSCGVYYPNEENPIYNNRESGYNDNDPHQINEQGNVGS
TSNANANANVTGNPSGAMAASSTNCFPFMFGDGDQNLWESYGLRLSKEENQ*

>G869 (428..1402)

AGGAACAGTGAAAGGTTCCGGTTTTTTGGGTTTCGATCTGATAATCAACAAGAAAAAGGG
TTTGATTATAGTCGGCTGGGTTTGAATCGACTGTGATTTGTCTTTGATTTCATATCTCTT
CTCCGATTTTCATCATCATCTTCCCCATCATCGTCGTCTTTGAAATCTTGTCTTCTCAACG
CTCTTCACTTCTGCTGTAATAAGCAGAGGCTTGTCTGGAGACTCCTTCTCTTCCATGC
GCTTAAGACCCAAAGGACTTGTCTAGTGTTGAAGTCTTTGGGGTTTTTCACATAAAGC
AGCAAAAGTTTTCTTTTTCATAGTTTCGCTGAGAGTTTGTAGTTTGTATACCAAAAAGT
TTTGACCTTTTAGAGTGATTTTTTGTCTTTCTGTTTTCTGGGTATTTTTGAGGAGTGGG
TTTAACAATGGTTGCGATTAGAAAGGAACAGTCTTTGAGTGGTGTAGTAGCGAGATTAA
GAAGAGAGCTAAGAGAAACACTCTATCGTCCCTTCTCAAGAAACCAACCTTTGAGGAA
AGTCCGTATTATTGTGAATGATCCTTATGCTACTGATGATTCTCTAGTGATGAGGAAGA
GCTTAAGGTTCTTAAGCAAGGAAAATGAAACGTATCGTTCGTGAGATTAACTTTCCTTC
TATGGAAGTTTCTGAACAGCCTTCTGAGAGTTCTTCTCAGGACAGTACTAAAAGTATGG
CAAGATAGCTGTGTGACGTTCTCTGCTGTTCTTAGGAAGAAGCCTGTTGGTGTTAGGCA
AAGGAAATGGGGGAAATGGGCTGCTGAGATTAGAGATCCTATTAAGAAAAGTAGGACTTG
GTTGGGTACTTTTGATACTCTTGAAGAAGCTGCTAAAGCTTATGATGCTAAGAAGCTTGA
GTTTGATGCTATTGTTGCTGGAATGTGTCCACTACTAAACGTGATGTTTCTTCATCTGA
GACTAGCCAATGCTCTCGTTCTTCACTGTTGTTCTGTTGAGCAAGATGACACTTCTGC
ATCAGCTCTCACTTGTGTCAACAACCTGATGACGCTCTGACCGTTGCTCCAAGTCTCC
AACTCCAATGTTCTGCTGCTGGTGGAAACAAGGAAACGTTGTTGATTTTCGACTTTACTAA
TCTACAGATCCCTGATTTTGGTTTCTTGGCAGAGGAGCAACAAGACCTAGACTTCGATTG
TTTCTCGCGGATGATCAGTTTGATGATTTCCGGCTTGCTTGATGACATTCAAGGATTCTGA
AGATAACGGTCCAAGTGCCTTACCAGATTTCGACTTTGCGGATGTTGAAGATCTTCAGCT
AGCTGACTCTAGTTTCGGTTTTCTTGATCAACTTGCTCCTATCAACATCTCTTGCCATT
AAAAAGTTTTCAGCTTCATAGGATCTTGCTTAGTAATGTTAAGTGAGAAGAGTGTTTTG
TTTTTTCGTTTATGCTTTAGTAATTAAGACATACAAAAGTGTGTGTTCCGGATTGTAGT
AAGATCTTAAGACATAAAGCCGGTTTTTGCAATTAGGAATCGAGTTTAAATGAAGTTTAA
GTTTATGTTTG

>G869 Amino Acid Sequence (domain in AA coordinates: 109-177)
MVAIRKQSLSGVSEIKKRAKRNLTSSLQPQETQLRKVRIIVNDPYATDDSSSDEELK
VPKPRMKRIVREINFPMSMEVSEQPSSESSQDSTKTGKIAVSASPAVPRKKPVGVRQRK
WGKWAABIRDPIKKTRTLGLTFDTLEEAAYDAKKLEFDAIVAGNVSTTKRDVSSSETS
QCSRSSPVVPVEQDDTSASALTGVNPDVSTVAPTPTPNVPAGGNKETLFDLFDFTNLQ
IPDFGFLAEQQDLDFDCFLADDQDFDLGLDDIQGFEDNGPSALPDFDFADVEDLQLAD
SSFGLDQLAPINISCLKSFAS*

>G1645 (25..1104)

CGTCGACCTCCCAACACTAACTCCATGTTTATAACGGAAAAACAAGTGTGGATGGATGAG
ATCGTCGCAAGAAGAGCTTCTTCTTCTTGGGACTTCCCTTTCAACGACATTAAATATTCAT
CAGCATCATCATCGTCACTGCAACACAAGTCATGAGTTTGAATCTTGAAGAGTCTCTT
GGAGATGTAGCGGTTACGAAGAAGAGAGTAATAATAAACCCTAATTTTCAGTAACAGC
GAGAGTGGTAAGAAGGAGACAACAGATAGTGGTCAGTCTTGGTCTCGTCGTCTTCAAAA
CCATCGGTCTTGGGGAGAGGACATTGGAGACCAGCTGAAGATGTTAACTCAAGAGACTT
GTCTCCATTACGGCCCAACAAAAGCTGGAACCTCATAGCTGAAAAGCTTCAAGGAAGATCT
GGGAAGAGCTGTAGACTACGATGGTTTAAACCAATTGGACCCGAGGATAAACCGAAGAGCT
TTCACAGAAGAAGAAGAGGAGAGGCTGATGCAAGCACATAGGCTTTATGGTAACAAATGG
GCAATGATTGCGAGGCTTTTCCCTGGTAGAACTGATAATTCAGTGAAGAACCATTGGCAT
GTTGTGATGGCTCCTAAGTATAGAGAACTCTTCTGCTTACCGTAGGAGAAAGCTTATG
AGTAATAATCCACTTAAACCTCACCTCACCAATAATCATCATCTTAACCTAACCTTAAT
TACCACTCTTTTATCTCCACTAATCACTTTCGCTCAGCCTTTCCCGAGTTTAAATTTG
ACTCATCACTGGTTTAAATAATGCCCTATCACGAGTGACCATAACCAGCTTGTGTGCTT
TTCCATTGCTTTCAAGGTTATGAGAACAATGAACCTCCGATGTTGTGAGTATGTTTGGC

>IGLC
>GL645 Amino Acid Sequence (domain in AA coordinates: 90-210)
MFITEKQVWMDEIVARRASSWDFPFNDINIHQHHHRHCNTSHEFEILKSP LGDVAVHEE
ESNNNNPNFSNSES GKKETDTS GQSWSSSSSKPSVLGRGHWRPAEDVKL KELVSIYGPQN
WNLIAEKLQGRSGKSCRLRFWNQLDPRINRRAFTEEEERLMQAHRLYGNKWAMIARLFP
GRTDNSVKNNHWHVVMARKYREHSSAYRRRKLMSNNPLKPHLTNNHHPNPNPNYGSFISTN
HYFAQPFPEPNLTHLNVNAPITSDHNQLVLPFHC FQGYENNEP MVVSMFGQMMVGDN
VGATSDALCNI PHIDPSNQKEPEPNDA MHWIGMDA VDEVEFEKAKQQPHFFDFLGLGTA*
>GI038 (240..1574)

>G1038 Amino Acid Sequence (domain in AA coordinates: 198-247)
MEKSGFSPVGLRVLVDDDP T W L K I L E K M L K K C S Y E V T T C G L A R E A L R L L R E R K D G Y D I V
I S D V N M P D M D G F K L L E H V G L E L D L P V I M M S V D G E T S R V M K G V H T G A C D Y L L K P I R M K E L K
I I W Q H V L R K K L Q E V R D I E G C G Y E G G A D W I T R Y D E A H F L G G G E D V S F G K K R K D F D E F K K L L
Q D E S D P S S S S K K A R V V W S F E L H H K F V N A V N Q I G C D H K A G P K K I L D L M N V P W L T R E N V A S
H L Q K Y R L Y L S R L E K G K E L K C Y S G G V K N A D S S P K D V E V N S G Y Q S P G R S S Y V F S G G N S L I Q K
A T E I D P K P L A S A S L S D P N T D V I M P P K T K T R I G F D P P I S S A F D S L L P W N D V P E V L E S K P
V L Y E N S F L Q Q Q P L P S Q S S Y V A I S A P S L M E E E M K P P Y E T P A G G S S V N A D E F L M P Q D K I P T V
T L Q D L D P S A M K L Q E F N T E G D S E E A *
>G1073 (62..874)

CCCCCGACCTGCCTCTACAGAGACCTGAAGATTCCAGAACCCACCTGATCAAAAATAA
CATGGAACCTAACAGATCTGAAGCAGACGAAGCAAAGGCCGAGACCACTCCCACCGGTGG
AGCCACCAGCTCAGCCACAGCCTCTGGCTCTTCTCCGGACGTCGTCCACGTGGTCTGTCC
TGCAGGTTCCAAAAACAAACCCAAACCTCCGACGATTATACTAGAGATAGTCTTAACGT
CCTTAGATCACACGTTCTTGAAGTCACCTCCGGTTCGGACATATCCGAGGCAGTCTCCAC
CTACGCCACTCGTCGGGCTGCGGCTTTGCATTATAAGCGGCACGGGTGCGGTCACTAA
CGTCACCAATACGGCAACCTGCGGCTCCGGCTGGTGGAGGTGTGATTACCGTCATGGTCCG
GTTTGACATTTTGTCTTTGACCGGTACTGCGCTTCCACCGCTGCACCACCGGGAGCAGG
AGGTTTGACGGTGTATCTAGCCGGAGGTCAAGGACAAGTTGTAGGAGGAATGTGGCTGG
TTCGTTAATTGCTTCCGGACCGGTAGTGTGATGGCTGCTTCTTTGCAAACGCAGTTTA
TGATAGGTTACCGATTGAAGAGGAAGAAACCCACCGCCGAGAACCACCGGGGTGCAGCA
GCAGCAGCCGGAGGCGTCTCAGTCGTGCGAGGTACGGGGAGTGGGGCCAGGCGTGTGA
GTCAAACCTCCAAGGTGGAATGGTGGAGGAGGTGTGCTTTCTACAATCTTGAATGAA
TATGAACCAATTTTCAATTTCTCCGGGGAGATATTACGGTATGAGCGGCGGTAGCGGAGG
AGGTGGTGGCGGTGCGACTAGACCCGCGTTTTAGAGTTTGTAGCGTTTGGTGACACCTTT
TGTTGCGTTTGCCTGTTTGACCTCAAACCTACTAGGCTACTAGCTATAGCGGTTGCGAAAT
GCGAATATTAGGTT

>G1073 Amino Acid Sequence (domain in AA coordinates: 33-42, 78-175)

MELNRSEADEAKAETPTGGATSSATASGSSSGRRPRGRPAGSKNPKPPTIITRDSNV
LRSHVLEVTSGSDISEAVSTYATRRGCGVCIISGTGAVTNVTIRQPAAPAGGGVITLHGR
FDILSLTGATLPPPPAGGLTVYLGGQGVVGGNVAGSLIASGPVVLMAASFANAVY
DRLPIEEETPPPRITGVQQQPEASQSSEVTGSGAQACESNLQGGNGGGVAFYNLGMN
MNNFQFSGGDIYMSGSGSGGGGATRPAP*

>G1146 (129..3095)

cttctctagcgtcactcttcttcttctcattggctcggtagaataaggccaaggaagggatca
gttttaagttttgtttcattcttcttctgtagtgagaaaaagagttttgaaaatcaaaac
aqaacaaaaatgcccatttaggcaaatgaaagatagctctgagactcacttagttatcaaaa
cccaacctttaaagcaccacaatccaaaaaccgttcaaaacggttaaatccctcctcctt
ctccttctccggtgacggtgactactccggcgacggttactcagagtcaagcttcttcac
cttcaccaccgtcaaaagaatcgtagccggaggagaaaccgtggtggaagaaaatctgatc
aaggagatgtttgtatgagacctagctctcgtcctcgtaaaccgccaccgccaagtcaaa
ccacttctccgcccgtctccgtcgccaccgcccgtgagattgtcgctgtgaatcatcaga
tgcagatgggtgttcgtaaaaactcaaaactttgctccaagacctggatttggaacacttg
gaactaaatgcattgttaaagctaaccactttctcgctgatttgacctaccaaggatttga
atcagtatgatgttacaattactcctgaagtgtcatcaaagagtggttaacagagctataa
ttgctgagtttagttagactttacaaagagctctgatctcgggaggagacttccggcttacg
atggccggaaaagtctttactctgctggagaacttcttttacttggaaggagttcagtg
ttaagattgttgatgaagatgacggtatcatcaatggccctaaaaggagagatcatata
aggtggcaatcaagttgttgacggggcaaatatgcatcacttaggcgagtttctagctg
gtaaacgggcagattgtccgcaagaggcggtgcagattcttgatattgtactcagggagt
tgtcgggttaagaggtttgtcccgttggaagatcttcttcttcgcctgatattaaaacac
cgcagcgactcgggtgaaggggttagagtcaggtgtggttttaccagagatttagaccaa
ctcaaatgggtttatcactaaatatcgatatggcttcagctgcattcatcgagcctcttc
cagtgatagagttttagcacagcttcttggaaggatgtcttgctgaagccattgtcgg
attctgatcgcgctcaagattaagaagggctcttagaggagtgaagtagaggttactcaca
gagcgaatgtaagaaggaaataccgtgttcgggtttaacaactcaaccaacaagagagc
taatgtttccagtagatgagaactgtactatgaagtcagttatttgagtatttccaagaga
tgtatggattcacgatccagcacacgcatttgccatgtctccaagttggaaccaaaga
aggcaagctatttgccgatggagcatgcaaaattgtcgagggaacacggtacacgaaaa
gggtgaatgagaagcagattactgctctcttgaaagttacatgccaaaggccgagggac
agagaaacgatatttgccgactgtccaacacacgcgatatgatcaagatccatattgcaa
aggagtttgcatgaacataagcgaaggttagcttctgttgaaagctcgattcttccag
ctccatggcttaagtatcacgagaacgggaagaaaaagattgtctcccgcaagttggtc
agtgaatatgatgaacaagaaaatgatcaacgggatgactgtgagcagatgggcctgtg
ttaacttctcacgcagcgttcaagaaaacgttgctcggtgattttgtaatgaactgggtc
agatgtgtgaagctcagggcatggagtttaatccagaacccgtgataccaatatatagtg
cgaggcccgatcaagtcgagaaaagctctaaagcatgtttatcacacttcaatgaacaaaa

ccaaaggcaaagagtttagagcttctgctggcaatattacctgataacaacgggttcacttt
atgggtgatcttaagagaatctgtgaaaccgagcttggtttgatattcctaatggtgtctca
caaaacatgtgttcaagatttagcaaacagtatctggcagatgtatcccttaaaatcaacg
taaagatgggaggaaggaacacagttcttagtagacgccataagctgtagaattccactgg
ttagcgataataccgacaatcatttttggcgagacgtgactcaccagagaacgggggaag
agtcaagcccttcaatcgctgctgtgtgtgcttctcaagactggcctgaagtgaacaaat
atgcgggttttagtttgctcaagctcacaggcaagaacttatacaagatttgtataaaa
catggcaagatcctgttcgcggtactgttagtggcggtatgatcagggaaccttcttatct
catttagaaaagcaacagggaacaaaccgcttcgaattatctttatcgtgatggagtaa
gcgaagggaattctatcaagttttactctatgagttggatgcaattcgaaaggcttggtg
catcgcttgaaacgaatttatcagccaccggtgacattcattagttgtacagaagcgtcacc
acactcgtttgttctgaataatcaccgagacaaaaacagtactgaccgaagcggaata
tcttaccaggtactgtagttgacactaaaaatgtcatccaactgaattcgacttctacc
ttttagtagcatgcgggtattcaggggaacagcaggcctgcacattaccatgttctttggg
acgagaacaatttcacagcagatggtattcaatctctgactaacaatctctgttatacct
atgcgcgggtgactcggtcggtctctatagttcctccagcgtattatgctcatcttgag
catttcgagcacgtttctacctggaacctgagataatgcaagacaacggatcacggggtg
aaaagaacacgaaacaaacaaactgtcggagacgtaggtgtgaagcctttaccagccttga
aggagaatgtgaagagagtaattgttctactgctaaaaatccaaacattccttaatacagtt
ttaataagtagtttgggtgttggctgttagttcggttttagatttaccatgtttttctt
atgtaaattttgcggttgggttaagccttttaggaattagtgattagggtttttctaa
agttgtacttttagctgatgataacgttgatgcagtgactttgttaaacctcctcttcta
cagtagtggttacgtcggttcctc

>G1146 Amino Acid Sequence (domain in AA coordinates: 886-896)

MPIRQMKDSSETHLVIKTQPLKHNPKTVQNGKIPPPSPSPVTVTTPATVTSQASSPSP
PSKNRSRRNRGRKSDQGDVCMRPSSRPKPPPSQTTSSAVSVATAGEIVAVNHQMOM
GVRKNSNFAPRPGFGLTGKTCIVKANHFLADLPKDLNQYDVTITPEVSSKSVNRALIAE
LVRLYKESDLGRRPLPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIINGPKRERSYKVA
IKFVARANMHHLGEFLAGKRADCPQEAQVILDIVLRELSVKRFCPVGRSFFSPDIKTPQR
LGEGLSWCGFYQSIRPTQMGSLNIDMASAAFIEPLPVIEFVAQLLGKDVLSKPLSDSD
RVKIKKGLRGVKEVTHRANVRKYRVAGLTTQPTRELMFPVDENCTMKSVIEYFQEMYG
FTIQHTHLPCLQVGNQKKASYLPMEACKIVEGQRYTKRLNEKQITALLKVTQRAEQRN
DILRTVQHNAYDQDPYAKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWN
MMNKKMINGMTVSRWACVNFSSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARP
DQVEKALKHVYHTSMNKTGKLELLAILPDNNGSLYGLDKRICETELGLISQCCILTKH
VFKISKQYLADVSLKINVKMGGRRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEES
PSIAAVVASQDWPEVTKYAGLVCAQAHQELIQDLYKTWQDPVRGTVSGGMIRDLLISFR
KATGQKPLRIIFYRDGVSEGFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTR
LFANNHRDKNSTDRGNILPGTVVDTKICHPTEFDLYLCSHAGIQGTSRPAHYHVLWDEN
NFTADGIQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRRARFYLEPEIMQDNGSPGKKN
TKTTTVGDVGKPLPALKENVKRMVFC*

>G1267 (152..967)

AAGTAGAGAATAATAATCACATCAAGATTGTTTATAACCCCTCCCNTAATCACCTTCTTA
NTNACCACCCTCTCCGGCTCTCAACAGAACAAACAAAAAACAGCTTCCGTTGTCCTG
TTCCGGCGAAATCGGACGGTCGAGATCAATCATGCATCGTAGAGCAGCAATTCAGAATC
GGATGACGAAGAAGATGAGACTTACAACGACGTCGTTCTGAAATCTCCTTCTTCTGTGA
AGACTCAAAGATCTCAAAACCAACTCCAAAGAAAAGGAGGAACGTAGAGAAGAGAGTTGT
CTCAGTTCCGATAGCTGACGTGGAAGGATCTAAGAGCAGAGGCGAAGTATATCCACCGTC
CGATTATGAGGCTGGAGAAAGTACGACAAAAACCGATCAAAGGCTCGCCTTATCCAG
GGGATATTACAGATGTAGTAGCTCAAAGGATGTCCGGCGAGGAAGCAGGTGGAGAGAAG
CCGTGTGGACCCCTTCTAAGCTTATGATTACTTACGCCTGCGACCACAATCACCTTTCCC
TTCTCTCCGCTAACACCAAATCCACACCGCTCCTCCGTCGTCCTCAAACCGCAAA
GAAAGAGGAAGAATACGAAGAGGAGGAAGAAGAACTAACCGTCACCGCCGAGAGGAACC
ACCGGCGGACTTGATCTAAGCCACGTAGACTACCGTTGCTATTAGGCGGCTGCTACAG
CGAAATCGGAGAGTTCGGGTGGTTCTACGACGCGTCGATCTCATCATCTGTTTCTTC
GAATTTCTCGACGTAACCTTAGAGAGAGGTTTTTTCAGTAGGCCAAGAGGAAGATGAGTC
TTTGTTCCGTGATCTCCGTGATTTACCTGATTGCGCCTCCGTGTTCCCGCTGGGACTGT

TGCGACGGAGGAGCAACATCGAAGATGTGATTTTGGCGCCATTCTTTCTGTGATAGTTC
TAGATGAGTTTGTGTGTGTAGCCAAAACCAAAGAAAAACACAATTTTTTTATTTTCC
ACTGTAAAGGTGTATCAATGGTGGATTCAATTTTTTAAAAA

>G1267 Amino Acid Sequence (domain in AA coordinates: 70-127)
MHRRAAIQESDDEDETYNDVVPESPSSCEDSKISKPTPKRRNVKRVVSVPIADVEGS
KSRGEVYPPSDSWAWRKYGQKPIKGSPPRGYYRCSSSKGCPARKQVERSRVDP SKL MIT
YACDHNHPFPSSSANTKSHRRSSVVLKTAKKEEYEEEEELTVTAEEPPAGLDLSDHVD
SPLLLGGCYSEIGEFWFYDASISSSSGSSNFLDVT LERGF SVGQEEDESLFGLDGLDLPD
CASVFRRGTVATEEQHRRCDFGAIPFCDSSR*

>G1269 (88..951)
AACAATTCTCTCTCTTTATTCTTCTTCTCAGCTTCAGATTTTCAGATCTTAAATCTTC
AAGTCTTCTTCTTCTTCTTCTGCAACCATGGCTATGCAGGAACGTTGTGAGAGTTTATGT
TCTGATGAACCTTATATCTTCTCAGATGCCCTTTACCTCAAGACAAGAAAGCCTTATACC
ATCACTAAACAAAGAGAGAGAAATGGACAGAGCAGAGCATGAGAAGTTTGTAGAAGCATTG
AAACTCTATGGCAGAGCTTGGAGACGAATCGAAGAACATGTTGGAACAAAACTGCAGTT
CAGATTCGAAGCCATGCGCAGAGTTCTTTACTAAGGTTGCTCGCGATTTTGGTGTAGC
TCTGAGTCCATTGAGATCCCGCCTCCAAGGCCAAAGAGAAAGCCGATGCATCCTTACCCT
AGAAAGCTTGTGATTCTGTATGCAAAAGAGATGGTATACGCTGAACCTAACCGATCCAAG
CTGATTTCAGGATGAAGATAACCGATCTCCAACATCGGTTTTATCAGCTCATGGCTCAGAT
GGATTAGGTTCCATTGGTTCAAATTCACCTAACTCTTCTCAGCTGAGTTATCATCTCAC
ACAGAGGAATCATTTCTCTCTAGAACGAGAGACCAACAGAGCCTTAAGCTCTTTGGAAAA
ACTTTTGTAGTTGGTGATTACAACCTTCAATGAGTTGTGATGATTCTGAAGATGGCAAG
AAGAAGCTATACTCAGAAACACAGTCTCTTCAATGTTCTTCTTCTACTTCAGAAAACGCT
GAAACAGAAGTGGTAGTGTGCGGAGTTCAAAGAAGTGAGAGATCAGCTTTCTCTCAGTTA
AAATCGTGGTGACTGAGATGAACAACATGAGAGGGTTTATGCTTACAAAAGAGAGTA
AAGGTGGAAGAAACATTGACAATGTAAATTTATCATATCCTTTGTGGTGAAGTGTTCGT
TTGTGTCAAGTCAGTTGTGTAACTCTTTTGATCTCAACATCAGATTATGTGTATAATGT
CAGAGTATTAGGGAAGTTTTTTTTGGATTAGATTTCGTAAGATCACTCCAAAGTTTCGTGT
CTTTCCATATAACCAAGTTAGAAATTGAGATCCTTGTACTTTAAACATTTTTATTGTATCAA
TCAAATCTTCTTGATGAAAAA

>G1269 Amino Acid Sequence (domain in AA coordinates: 27-83)
MAMQERCESLCSDELISSSDAFYLKTRKPYTITKQREKWTEAEHEKFVEALKLYGRWRR
IEEHVGTKTAVQIRSHAQKFFTKVARDFGVSSSEIEIPPPRPKRKPMHPYPRKLVIPDAK
EMVYAEITGSKLIQDEDNRSPTS VLSAHGSDGLSGISNSPNSSS AELSSTESLSLEA
ETKQSLKLFKFTFVVDYNSSMSCDSDGKKKLYSETQSLQCSSSTSENAETEVVVSEF
KRSERSAFSQLKSSVTEMNMRGFMPYKKRVKVEENIDNVKLSYPLW*

>G1452 (175..1296)
ATTTATTAAGCATCAATGAGAGAACTTCAGAGCTGGGTTTGAGTTCTGTCCAATAATACA
TAACCACGTTATCATTTTTGTCTTTACTATCTCATTACACTCTTCTGTTATTCGCCCAA
TTCTTACAGTCATTACTCTCTATAGGGCTCGAGCGGCCCGCCGGCAGGTTTCTATGCAG
ATGTTTCACACTTCCCGCTCCATTGCCAGATTGGGTTTCGGTTAAGTCGCAATTAGTA
CTCACTATAGGGCTCGAGCGGCCCGCCGGCAGGTAAAAGATCAAACAATGTCTAAAGAA
GCTGAGATGTCGATCGCGGTGTCGGCTTTGTTCCCTGGTTTTAGATTCTCTCTACTGAT
GTTGAACCTTATCTCGTACTATCTTCGTGTAATCGATGGTGATGAGAACTCTGTTGCT
GTGATTGCTGAGGTCGAGATTTACAAGTTTCGAGCCGTGGGACTTGCCAGAGGAATCGAAA
CTGAAATCGGAGAACGAGTGGTTTTACTTCTGCGCGAGGGGGAGGAAGTACCCGCACGGG
TCACAAAGCCGGCGAGCCACACAGCTAGGATATTGGAAGCGACCGGTAAGAGCGGAGT
GTTAAATCCGGGAACCAAGTTGTTGGAACCAAGAGAACGCTTGATTTTCATATCGGTGCG
GCTCCTCGTGGCGAGAGAACGGAGTGGATTATGCATGAATACTGCATCCATGGAGCCCCA
CAGGATGCATTAGTGGTGTGCCGTTAAGAAAAATGCTGATTTTGGGCTAGTTCGACC
CAAAAAATTGAGGATGGTGTGTGCAAGACGATGGCTACGTTGGCCAAAGAGGTGGTTTG
GACAAGGAGGACAAATCCTACTATGAATCTGAGCATCAGATACCAATGGTGACATCGCA
GAATCATCAAATGTTGTTGAGGATCAGGCCGATACCGATGATGATTGTTACGCCGAGATT
CTGAACGATGATATAATAAGCTCGACGAAGAAGCGTTGAAAGCTAGCCAAGCGTTTCGA
CCAACATAATCCAACCTCATCAAGAAACAATATCAAGCGAGTCATCGAGTAAGAGGTCAAAA
TGTGGTATAAAAAAGAATCAACGGAAACAATGAATTGTTACGCTTTGTTTCAGGATCAAG
AACGTTGCCGGAACCGACTCCAGCTGGAGATTCCCGAACCCGTTCAAATCAAGAAAGAT

GATAGCCAGAGATTGATGAAGAATGTTCTGGCCACTACTGTTTCTTGGCTATCTTATTT
TCTTTCTTTTGGACTGTATTAATAGCTAGGAAGCTAAAGCTAGTTACGACATACATATTAT
TTATACATAAATAAATATAGTATTTTGTCTATGGCAAAAAAAAAAAAAAAAAA

>G1452 Amino Acid Sequence (domain in AA coordinates: 30-177)
MQMVHTSRISIAQIGFGVKSQVLVTIGLERPPGQVKDQMSKEAEMSIASVSAFPGRFRSP
TDVELISYLLRRKIDGDENSVAIVAEVEIYKFEPWDLPEESKLKSENEWFFYCARGRKYP
HGSQSRRTQLGYWKATGKERSVKSGNQVVGTKRTL VFHIGRAPRGERTEWIMHEYCIHG
APQDALVVCRLRNADFRASSTQKIEDGVVQDDGYVGQRGGDLKEDKSYESEHQIPNGD
IAESSNVVEDQADTDDCYAEILNDDIIKLDEEALKASQAFRPTNPTHQETISSESSSKR
SKCGIKKESTETMNCYALFRIKNVAGTDSWRFPNPFKIKKDDSQRLMKNVLATTVFLAI
LFSFFWTVLIARN*

>G1494 (114..1406)
TCGACAGAGTTGTGTTGGGCGTGGAACCTGGACTAGTTCCACATATCAGGTTATATAGAT
CTTCTCTTTCAACTTCTGATTTCGTCCAGAAGCTTTCCTAATCTGAGATCTGACATGGAAC
ACCAAGGTTGGAGTTTGGAGGAGAATTATAGTTTGTCCACTAATAGAAGATCTATCAGGC
CACAAGATGAACCTAGTGGAGTTATTATGGCGAGATGGACAAGTGGTTCTGCAGAGCCAAA
CTCATAGAGAACAAACCCAAACCCAGAAACAAGATCATCATGAAGAAGCCCTAAGATCCA
GCACCTTTCTTGAAGATCAAGAACTGTCTCTTGGATCCAATACCCCTCCAGATGAAGACC
CATTCGAACCCGACGACTTCTCTCCACTTCTTCTCAACCATGGATCCCTCCAGAGAC
CAACCTCAGAGACGGTTAAGCCTAAGTCCAGTCCCTGAACCTCCTCAAGTCATGGTTAAGC
CTAAGGCCCTGTCTGACCCTCCTCCTCAAGTCATGCCTCCTCCAAAATTTAGGTTAACAA
ATTTCATCATCGGGGATTAGGGAACAGAAATGGAACAGTACTCGGTAACGACCGTTGGAC
CTAGCCATTGCGGAAGCAACCCATCACAGAACGATCTCGATGTCTCAATGAGTCATGATC
GAAGCAAAAACATAGAAGAAAAGCTTAATCCGAACGCAAGTTCCCTCATCAGGTGGCTCCT
CTGGTTGCAGCTTTGGCAAAGATATCAAAGAAATGGCTAGTGAAGATGCATCACAAACCG
ACCGTAAGAGAAAACGTATAAATCACACTGACGAATCTGTATCTCTATCAGATGCAATCG
GTAACAAGTCGAACCAACGATCAGGATCAAACCGAAGGAGTCGAGCAGCTGAAGTTTCATA
ATCTCTCCGAAAGGAGGAGGAGATAGGATCAATGAGAGAATGAAGGCTTTGCAAGAAC
TAATACCTCACTGCAGTAAACTGATAAAGCTTCGATTTTAGACGAAGCCATAGATTATT
TGAAATCACTTCAGTTACAGCTTCAAGTGATGTGGATGGGAGTGGAATGGCGGCGGCGG
CGGCTTCGGCTCCGATGATGTTCCCGGAGTTCAACCTCAGCAGTTTCATACGTGATAC
AGAGCCCGGTACAGTTACCTCGATTTCCGGTTATGGATCAGTCTGCAATTCAGAACAAATC
CCGGTTTAGTTTGCCAAAACCCGGTACAAAACAGATCATCTCCGACCGGTTTGCTAGAT
ACATCGGTGGGTTCCACACATGCAGGCCGCGACTCAGATGCAGCCGATGGAGATGTTGA
GATTTAGTTACCGGCGGGACAGCAAAGTCAACAACCGTCTGTGTGCCGACGAAGACCA
CCGACGGTTCTCGTTTGACCACTAGGTTGGTGAGCCACTTTGC

>G1494 Amino Acid Sequence (domain in aa coordinates: 261-311)
MEHQWSFEENYSLSTNRRSIRPQDELVELLWRDQVVLQSQTTHREQTQTKQDHHEEAL
RSSTFLEDQETVSWIQYPPDEDPFEPDDFSSSHFFSTMDPLQRPTSETVKPKSSPEPPQVM
VKPKACPDPPPQVMPPPKFRLTNSSSGIRETEMEQYSVTVGPSHCGSNPSQNDLDVSMS
HDRSKNIEEKLNPANSSSSGSSGCSFGKDIKEMASGRCTTDRKRKRINHTEDESVSLS
AIGNKSNQSRSGSNRRSRAAEVHNLSERRRRDRINERMKALQELIPHCSKTDKASILDEAI
DYLKSLQLQLQVMWMSGMAAAAAAPMMFPVQPPQFIRQIQSPVQLPRFPVMDQSAIQ
NNPGLVCQNPVQNIISDRFARYIGGFPHMQAATQMQPMEMLRFS SPAGQQSQQPSSVPT
KTTDGSRLDH*

>G1548 (1..2511)
ATGGCAATGTCTTGCAAGGATGGTAAGTTGGGATGTTTGGATAATGGGAAGTATGTGAGG
TATACACCTGAACAAGTTGAAGCACTTGAGAGGCTTTATCATGACTGTCTAAACCGAGT
TCTATTGCGCGTACAGAGTTGATCAGAGAGTGTCTATTCTCTTAACATTGAGCCTAAA
CAGATCAAAGTGTGTTTTCAGAACCGAAGATGTAGAGAGAAAACAAAGGAAAGAGGCTTCA
CGGCTTCAAGCTGTGAATCGGAAGTTGACGGCAATGAACAAGCTCTTGATGGAGGAGAAT
GACAGTTGACAGAAGCAAGTGTACAGCTGGTCCATGAAAACAGCTACTTCCGTCAACAT
ACTCCAAATCCTTCACTCCAGCTAAAGACACAAGCTGTGAATCGGTGGTGACGAGTGGT
CAGCACCATTGGCATCTCAAAATCCTCAGAGAGATGCTAGTCTGCAGGACTTTTGTCC
ATTGCAGAAGAACTTTAGCAGAGTTTCTTTCAAAGGCAACTGGAACCGCTGTGAGTGG
GTTTCAGATGCCTGGAATGAAGCCTGGTCCGATTCATTGGAATCATCGCTATTTCTCAT
GGTTGCACTGGTGTGGCAGCAGCGCCTGTGGCCTAGTGGGTCTTGAGCCTACAAGGGTT

GCAGAGATTGTCAAGGATCGTCCTTCGTGGTTCCGCGAATGTCGAGCTGTTGAAGTTATG
AACGTGTTGCCAACTGCCAATGGTGGAACCGTTGAGCTGCTTTATATGCAGCTCTATGCA
CCAACCTACATTGGCCCCACCACGCGATTCTTGGCTGTTACGTTACACCTCTGTTTTAGAA
GATGGCAGCCTTGTGGTGTGCGAGAGATCTCTTAAGAGCACTCAAAATGGTCCTAGTATG
CCACTGGTTTCAAAATTTGTGAGAGCAGAGATGCTTTCCAGTGGGTACTTGATACGGCCT
TGTGATGGTGGTGGCTCAATCATACATAGTGGATCATATGGATTTGGAGGCTTGTAGC
GTGCCTGAGGTCCTTGGCGCCCGTCTATGAGTCACCCAAAGTACTTGACAGAGAACAACA
ATGGCGGCACATGCGTCAGCTCAAGCAAATAGCTCAGGAGGTTACTCAGACTAATAGTAGT
GTTAATGGGTGGGACGGCGTCTGCTGCCCTTAAGAGCTCTCAGCCAGAGGCTAAGCAGA
GGCTTCAATGAAGCTGTAAATGGTTTCACTGATGAAGGATGCTCAGTGATAGGAGATAGC
ATGGATGATGTCAATCACTGTAAACTCTTCTCCAGACAAGCTAATGGGTCTAAATCTT
ACATTTGCCAATGGCTTTGCTCCTGTAAAGCAATGTTGTTTTATGCGCAAAAGCATCAATG
CTTTTACAGAATGTTCTCCGCGCATCTGCTTCGGTTTCTGAGGGAGCATAGGTCAGAA
TGGGCTGACAACAACATTTGATGCGTATCTAGCAGCAGCAGTTAAAGTAGGGCCTTGTAGT
GCCCCGAGTTGGAGGATTTGGAGGGCAGGTTATACTTCCACTTGCTCATACTATTGAGCAT
GAAGAGTTTATGGAAGTCATCAAATTGGAAGGTCTTGGTCATTCCCTGAAGATGCAATC
GTTCCAAGAGATATCTTCTTCTCAACTTTGTAGCGGAATGGATGAAAATGCTGTAGGA
ACCTGTGCGGAACCTATATTTGCTCCAATCGATGCTTCGTTGCGGATGATGCACCTCTG
CTTCTTCTGTTTTCGTATTATCCCTCTTGATTCGCGAAAGGAAGTATCTAGCCCCAAC
CGAACCTTGGATCTGCTTCGGCACTGGAAATGGTTTCACTGGAACAAAAGCCTCAACT
GATCAATCAGGAACTCCACATGTGCAAGATCTGTGATGACAATAGCATTGAGTTTGGT
ATCGAGAGCCATATGCAAGAACATGTAGCATCCATGGCTAGGCAGTATGTTGAGGTATC
ATATCATCGGTGCAGAGAGTAGCATTGGCTCTTCTCTCTCATATCAGCTCACAAGTT
GGTCTACGCACTCTTTGGGTACTCCTGAAGCCCAAACACTGCTCGTTGGATTTGCCAG
AGTTACAGGGGCTACATGGGTGTTGAGCTACTTAAATCAAACAGTGACGGAATGAATCT
ATTCTTAAGAATCTTTGGCATCACACTGATGCTATAATCTGCTGCTCAATGAAGGCCCTG
CCCGTCTTACATTGCAAACCAGGCGGACTTGACATGCTGGAGACTACATTAGTTGCT
CTTCAAGACATCTCTTTAGAGAAGATATTTGATGACAATGGAAGAAAGACTCTTTGCTCT
GAGTTCCACAGATCATGCAACAGGGCTTCGCGTGCCTTCAAGGCGGGATATGTCTCTCA
AGCATGGGGAGACCAGTTTTCGTATGAGAGAGCAGTTGCTTGGAAAGTACTCAATGAAGAA
GAAAATGCTCATTGCATCTGCTTTGTGTTTCAATGGTCTTTGTGTGA

>G1548 Amino Acid Sequence (domain in AA coordinates: 17-77)
MAMSKDGLKGLDNGKYVRYTPEQVEALERLYHDCPKPSSIRRQQLIRECPILSNIEPK
QIKVWFQNRRCREKQKREASRLQAVNRKLTAMNKLMEENDRLQKQVSQLVHENSYFRQH
TPNPSLPARRKDTSCSVVTSQHQQLASQNPQRDASPAGLLSIAEETLAEFLSKATGTAVEW
VQMPGMKPGPDSIGIIAISHGCTGVAARACGLVGLPRTVAEIVKDRPSWFRECRAREVM
NVLPTANGGTVELLYMQLYAPTTLAPPRDFWLLRYTSVLEDGSLVVCERSLKSTQNGPSM
PLVQNFVRAEMLSSGYLIRPCDGGGSIIHIVDHDLEACSVPEVLRPLYESPKVLAQKTT
MAALRQLKQIAQEVTQTNSSVNGWRRPAAALRALSQLSRGFNEAVNGFTDEGWSVIGDS
MDDVTITVNSSPDKMLGNLTFANGFAPVSNVVLCAKASMLLQNVPPAILLRFLREHRSE
WADNNIDAYLAAAVKVGPCSRVGGFGGQVILPLAHTIEHEEFMEVIKLEGLGHSPEDAI
VPRDIFLLQLCSGMDENAVGTCAELIFAPIDASFADDAPLLPSGFRIIPLDSAKEYSSPN
RTLDLASALEIGSAGTKASTDQSGNSTCARSVMTIAFEFGIESHMQEHVASMARQYVRGI
ISSVQRVALALSPSHISSQVGLRTPLTPEAQTARWICQSYRGYMGVELLKSNSDGNES
ILKNLWHHTDAIICCSMKALPVFTFANQAGLDMLETTLVALQDISLEKIFDDNGRKTLC
EFPQIMQOGFACLOGGICLSSMRPVSYERAVAWKVLNEENAHICFVFINWSFV*

>G1574 (1..1962)

ATGGATGATACAATGGACATGAGTTCAGGTAGTGATGAAGAAGTACAAGAAGAGAAGACC
ACTGTTAACGAGAGGGTCACTATCAGGCTGCATTACAAGATCTGAAGCAACCCAGACC
GAAAAGGATCTACCTCCTGGTGTCTTACAGTTCTCTTATGAGGCATCAGAAAATTGCA
TTGAACTGGATGCGTAAGAAAGAAAAAGAACAGGCAGTCTTTGGGAGGGATATTAGCA
GATGATCAGGGACTTGGTAAACGATCTCGACGATCTCTCTTATCCTGTTACAAAAGTTG
AAGTCACAATCAAAGCAGAGAAAGCGAAAAGGTCAAAACTCTGGTGGTACATTGATTGTT
TGTCACAGCAAGTGTGTAAACAATGGGCAAGAGAAGTTAAAGAGAAGGTTTCTGATGAA
CACAAACTCTCTGTTTGTAGTCCACCATGGATCTCACAGAACCAAGATCCAACAGAAATA
GCAATATATGATGTGGTCATGACAACTTACGCCATTGTTACAAATGAAGTTCCACAAAAC
CCTATGCTGAATCGTTATGATAGTATGAGAGGCAGAGAAAGCCTTGACGGATCGAGTTTG

ATTACAGCTCAGCTGGTGCCTAGGAAGAGTTAGGTGGTTGAGAGTAGTATTAGATGAA
GCTCATACAATTAAAAACCATAGAACCTTAATTGCAAAAGCTTGTTTTAGCCTTAGAGCC
AAAAGGAGATGGTGTGTTGACTGGAACGCCGATAAAGAACAAAGTAGACGATCTTTATAGC
TATTTTCAAGATTTCTTAGATATCATCCATATGCCATGTGCAATTCATTTACCAAAGAATC
AAAGCTCCAATTGATAAAAGCCTCTTCATGGTTACAAGAAGCTTCAAGCTATTCTAAGG
GGTATAATGTTGCGCCGACCAAGAATGGTCTTTCTACAGGAAGCTTGAATTGAATTCA
CGTTGGAAGTTTGAGGAATATGCTGCTGATGGGACTTTGCATGAACACATGGCTTATCTT
TTGGTGATGCTTTTTCGACTACGCCAAGCTTGTAAACCATCCACAACCTTGTAAACGGATAT
AGTCACTCAGATACTACAAGAAAAATGTCAGATGGAGTTCGAGTAGCCCCCTAGAGAGAAT
CTAATCATGTTCCCTCGATCTCTTGAAATTATCCTCAACCACCTGCTCTGTTTGTAGTGAT
CCACCAAAAGACCTGTTGTTACTTTGTGTGGCCATGTGTTTTGTATGAGTGTGTGTCT
GTAAACATTAAACGGGGATAACAATACGTGCCCTGCACCTTAATTGCCACAGCCAGCTTAAA
CATGATGTTGTTTTCACTGAATCTGCAGTTAGAAGTTGCATCAACGATTATGATGATCCT
GAAGATAAAATGCTTTAGTTGCATCAAGGCGAGTTTATTTTCATCGAAAATCCGAGCTGT
GATAGAGATTTTTCGACTTGCAGAGCAAGGCAGTCCAGACACTCCACCAATAAAGAC
AATAGTATCAGTGGACTGAATCTCATTTTTACGTTTCTCAAAGACAAATGTAATGATTAT
GAAACAGGTGCGATGTTGATGTCTCTTAAAGCTGGAAACCTTGGATTGAATATGGTAGCT
GCAAGTCATGTCTTCTACTGGACCTATGGTGAATCCAACAACAGAGGATCAAGCTATT
GATCGAGCTCATCGTATCGGACAAACTCGAGCTGTACGGTCACCTCGTATTGCCATCAA
AATACCGTTGAGGAACGAATTTTGAAGCTCTTCATGAACGTAAAAGGAACATTGTTGCATCT
GCATTGGGTGAAAAAACTGGCAAAAGTTCTGCGATTCAACTAACACTAGAAGATCTCGA
ATATCTGTTTTTTTGGTGTGTAGAATATCCAGAGTTTTTATTGATAAGAGGAATAAAACC
TTTAGCTATTTAATAAGTCACAAGTGTGAATGTAATGAATAA

>G1574 Amino Acid Sequence (domain in AA coordinates: 28-350)
MDDTMDSSGSDEEVQBEKTTVNERVIYQALQDLKQPKTEKDLPPGVLTVPMLMRHQKIA
LNWMRKKEKRSRHLGILADDQGLGKTIISTISLILLQKLKSQSQQRKRKGQNSGGTLIV
CPASVVKQWAREVKEKVSDEHKLKSLVHGHSHRTKDPTEIAIYDVVMTTYAIVTNEVPQN
PMLNRYDSMRGRESLDGSSLIQPHVGALGRVRLRVVLDEAHTIKNHRTLIAKACFSRLA
KRRWCLTGTPIKNKVDDLYSYFRFLRYHPYAMCNSFHQRIKAPIDKKPLHGYKKLQAILR
GIMLRRTKEWSFYRKLELNSRWKFEEYAADGTLHEHMAYLLVMLLRRLRQACNHPQLVNGY
SHSDTTRKMSDGVVRVAPRENLMFLDLLKLSSTTCSVCSDDPPKDPVVTLCGHVFCYECVS
VNINGDNNTCPALNCHSQLKHDVVFTESAVRSCINDYDDPEDKNALVASRRVYFIENPSC
DRDSSVACRARQSRHSTNKNDSISGLNLI FTFLKDKCNDYETGAMLSLKAGNLGLNMVA
ASHVILLDLWNPTTEDQAIDRAHRIGQTRAVTVTRIAIKNTVEERILTLHERKRNIVAS
ALGEKNWQKFCDSINTRRSRISVFWCWEYPRVFIDKRNKTFSYLISHKKECNE*

>G1586 (1..807)

ATGAATCAAGAAGGTGCTTCACATAGCCCATCCTCCACTTCCACCGAACCAGTCCGGGCA
CGTTGGTCACTAAACCGGAGCAAATCTTGATACTCGAATCCATCTTCAACAGTGGTACT
GTTAACCCACCAAAAGATGAAACGGTGAGGATAAGAAAGATGCTTGAGAAATTCGGTGCT
GTGGGAGACGCAACGCTCTTCTACTGGTTTCAAACCGACGGTCAAGATCTCGCCGGAGA
CACCGGCAGCTTTTAGCAGCCACCACCGCAGCCGCCACCTCCATAGGAGCTGAAGACCAC
CAGCACATGACGGCCATGAGCATGCATCAATATCCTTGCAGCAACAACGAGATTGATTTG
GGGTTTGGAAGTTGTAGCAACTTATCAGCTAATTACTTCTTAAATGGATCGTCGTCATCT
CAAATCCCTTCTTTTCTCTCGGCCTCTCTTCTTCAAGTGGTGGGTGTGAGAACAACAAT
GGTATGGAGAATCTCTTCAAATGTATGGCCATGAATCTGATCATAATCATCAGCAGCAG
CATCATAGCTCAAATGCTGCATCAGTTTAAACCCATCTGATCAAAACTCCAACCTCCAA
TACGAACAAGAAGGGTTTATGACGGTGTATATAACGGAGTTCCTATGGAAGTAACAAA
GGAGCAATAGACATGAAAAAATGTTCCGGTGATGATTCCGGTGTACTTCATTCTCTGGT
CTTCTCTTCCCACTGATGAGTTTGGTTTCTTGTATGCATTCTTTACAACATGGACAAACT
TATTTCTGTTACCGAGACAGACATGA

>G1586 Amino Acid Sequence (domain in AA coordinates: 21-81)
MNQEGASHSPSSTSTEPVRARWSPKPEQILILESIFNSGTVPNPKDETVRIRKMLEKFGA
VGDANVFYWFQNRRSRRRRHQLAATTAAATSIGAEDHQHMTAMSMHQYPCSNNEIDL
GFGSCSNLSANYFLNGSSSSQIPSFLLGLSSSSGCGENNGMENLFKMYGHESDHNHQQQ
HHSSNAASVLNPSDQNSNSQYEQEGFMTVFINGVPEVTKGAIDMKTMFGDDSVLLHSSG
LPLPTDEFGFLMHSLOHGQTYFLVPRQT*

>G1786 (1..1170)

ATGATCGTGACGGTGGGGGAGCATCCGAGGACGGTGAAGGTGGAGGGGTGGTTCTGAAG
AAAGGGCCATGGACGGTGGCCGAGGACGAGACACTGGCGGCTTACGTACGGGAATACGGT
GAAGGGAACTGGAATTCTGTTTCAAGAAGACATGGCTGGCTAGGTGTGGCAAGAGCTGC
CGCCTCCGCTGGGCTAACCACTTACGACCTAATCTCAGGAAAGGCTCCTTACCCCCGAG
GAAGAAGCTCTCATCATACAACCTCCACTCTCAGCTAGGCAACAAATGGGCTCGCATGGCT
GCTCAGTTACCAGGCAGAACAGATAACGAGATCAAGAACTACTGGAACACGAGGTTGAAA
CGCTTCCAACGCCAAGGCCTCCCTCTCTACCTCCAGAATATTTCCAAAACAATCATCAA
CAACAAATGTATCCTCAACAGCCCTCCTCACCTCTCCGTCCTCAACACCTGCTTCTTCC
TTTACCTTTCTCTCCCAACCGCCTTCTCTGTGTCCCAACGTTGTTATAACACTGCC
TTCTCTCCCAAGGCCTCATATATTTCTCTCTCAACCAATTTCTTGTCTCGTCTCCGACC
TTTCTTACACCCATTCTCTCTTTCTCTCTATCAGTCTACCAATCCGGTTTACTCCATG
AAACATGAGCTCTCTTCAACCAAAATTCATACCTCTGCTCTTTAGGAGTCTATCAAGTA
AGCAAGTTCTCAGACAATGGGGATTGTAACCAAAACCTGAACACCGGTTTGCATACAAAT
ACCTGTGAGCTGTTAGAGGATCTTATGGAGGAGGCCGAGGCTCTAGCTGATAGCTTTCTGT
GCTCTTAAGCGGAGACAAATCATGGCTGCGCTTGAGGACAACAACAACAACAACCTTT
TTCTCGGGAGGTTTTCGGGACATCGTGTCTTCTTCAACAGTCTATGTTCTTGCAGGTTTA
ACACCAAGGAAGATGAGTCTCTCCAGATGAACACAATGCAAGATGAGGACATAACAAAG
CTTCTTGACTGGGGAAGTGAAGTGAAGAAATCTCAACGGGCAATCCTCTGTGATAACA
ACAGAGAACAACCTTGTCTTGACGATCACCAGTTCGCTTTTCTGTTTCCAGTTGATGAT
GACACCAACAACCTTGCCAGGGATCTGCTAG

>G1786 Amino Acid Sequence (domain in AA coordinates: TBD)
MIVYGGGASEDGEVGGVVLKGPWTVAEDETLAAYVREYGEENWNSVQKKTWLRACGKSC
RLRWANHLRLNLRKGSFTPEERLIIQLHSQLGKWKARMAAQLPGRITDNEIKNYWNTRLK
RFQRQGLPLYPPEYSQNNHQQMYPQPPSSPLPSQTPASSFTFPLLPQPSLCPKRCYNTA
FSPKASYISSPTNFLVSSPTFLHTHSSLSYQSTNPVYSMKHELSSNQIPYSASLGVYQV
SKFSDNGDCNQNLNTGLHTNTCQLLEDLMEEAEALADSFRAKRRQIMAALEDNNNNNNF
FSGGFGRVSSNSLCSLQGLTPKEDESLOMNTMQDEDITKLLDWGSESEEBISNGQSSVIT
TENNLVLDDHQFAFLFPVDDDTNNLPGIC*

>G1792 (77..496)

AATCCATAGATCTCTTATTAATAACAGTGCTGACCAAGCTCTTACAAAGCAAACCAATC
TAGAACACCAAAGTTAATGGAGAGCTCAAACAGGAGCAGCAACAACCAATCACAAGATGA
CAAGCAAGCTCGTTTCCGGGGAGTTTCAAGAAGGCCTTGGGGAAAGTTTGCAGCAGAGAT
TCGAGACCCGTCGAGAAACGGTGCCCGTCTTTGGCTCGGGACATTTGAGACCGCTGAGGA
GGCAGCAAGGGCTTATGACCGAGCAGCCTTTAACCTTAGGGGTCTATCTCGCTATACTCAA
CTTCCCTAATGAGTATTATCCACGTATGAGCAGTACTCTGCTTCGCCCTCCTTATGCTTC
TTCTTCTTCTGTCGTCGTCATCGGGTTCACTTCTACTAATGTGAGTCGACAAAACCAAG
AGAAGTTTTCGAGTTTGTATTTGGACGATAAGGTTCTTGAAGAACTTCTTGATTGAGA
AGAAAGGAAGAGATAATCACGATTAGTTTTGTTTTGATATTTTATGTGGCACTGTTGTGG
CTACCTACGTGATTATGTGCATGTATAGGTGCTTGTATTAGTACTTTATAACATGCATG
CCACGACCATAAATTGTAAGAGAAGACGTACTTTGCGTTTTTCATGAAATATGAATGTTAG
ATGGTTTGAGTACAAAAA

>G1792 Amino Acid Sequence (domain in aa coordinates: 17-85)
MESSNRSSNNQSQDDKQARFRGVRRRPWGKFAAEIRDPSRNGARLWLGTFFETAEAAARAY
DRAAFNLRGHLAILNFPNEYPRMDYSLRPPYASSSSSSSGSTSTNVSRQNRQEVFEF
EYLDDKVLEBLDSEERKR*

>G1865 (48..899)

AAGAAGAGGACATGAAGCACAGAGATTCTGCAGACTGCAGGTGACCAATGGACACTTTAT
CAATAAAACATACCTACTACTCTCTTACACTTTCAATTTTCCAATACAAATCCCAATCT
TTAATCTCTCTTCTTCTTCTCTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCTCTTCT
CATTTACAGAAATCACAATGGGAAGAACTTGAAGAACCAAGCTCTTGTGTTCAAGTACTTAG
CTGCAATATGCCTGTTCCACCTCATCTTCTCTTCTCTCATCAAAAGACCCCTTCTCTTCT
CTTCT
TTGGGTGGAATGTGTATGAGATGGGAATGGGAAGAAAGATAGATGCAGAGCCAGGAAGAT
GTAGAAGAACTGATGGCAAGAAATGGAGATGCTCTAAGAAAGCTTACCCTGACTCTAAGT
ACTGTGAGAGACATGATGATAGAGGCAAGAACCGTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
CTCAATTCACCTCAAAATCTCTTCTCGACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT
TGGATGATTTCTTCTCATAGAACCTTCCGGGTCAATCAAAGCTGCTCTGGCTCAGCAA

TGGAAGATAATGATGATGGCTCATGTAGAGGCATCAACAACGAGGAGAAGCAGCCGGATC
GACATTGCTTCATCCTTGGTACTGACTTGAGGACACGTGAGAGGCCATTGATGTTAGAGG
AGAAGCTGAAACAAAGAGATCATGATAATGAAGAAGAGCAAGGAAGCAAGAGGTTTATA
GGTTTCTTGATGAATGGCCTTCTTCTAAATCTTCTGTTTCTACTTCACTCTTCATTTGAT
CATCTTTTGTCTTATAACCTTGTATTTCTTGTTAAGATGGTAATGCAAATT
>G1865 Amino Acid Sequence (domain in AA coordinates: 124-149)
MDTLSIKTYLLLSYTFNFIQIPFNLSFFFISLSLSLFMATRIPFTESQWEELENQALV
FKYLAANMPVPPHLLFLIKRPFLFSSSSSSSSSSSFFSPTLSPHFGWNVYEMGMGRKIDA
EPGRCRRTDGKKWRCSEAYPDSKYCERHMRGKNRSSSRKPPPTQFTPNLFLDSSSRRR
RSGYMDFFSIEPSGSIKSCSGSAMEDNDDGSCRGINNEEKQPDHCFILGTDLRLTRERP
LMLEEKLKQRDHDNEEQGSKRFYRFLDEWPSSKSSVSTSLFI*
>G1886 (43..909)
AGGAAACATAAGTAATCGTTGCTTCGATCCTTTGTACATGGATGGATCCTGAACAGGAA
ATCTCAAACGAGACTTTGGAACTATATTGGTAAGTTCAACAAAAGGAAGCAATAATAAC
AATAAGAAAATGGAAGAAGAAATGAAGAAGAAAGTATCAAGAGGAGAATTAGGAGGTGAA
GCTCAAAATTGTCCAAGATGTGAATCTCCAAACACAAAGTTTTGTACTACAACAATAT
AGTCTCTCACAACTCGTTACTTCTGCAAATCTTGTGCGAGATATGGACTAAAGCGGT
ACTCTTCGTAACGTTCCCGTCGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
GCTTTCTCCAAGAACAACAATAAGTCTATTAATTTCCATACTGATCCACTTCAGAAC
CCTTTAATTACGGGAATGCCACCATCATCTTTTGGTTATGATCACTCCATTGATCTCAAC
CTCGCTTCGCTACTCTCCAAAAGCATCATTTATCCTCTCAAGCTACTACGCCTTCTTTT
GGGTTTGGAGGTGATCTTTCTATTTATGGAACCTCAACGAATGATGTAGGGATCTTCGGA
GGGCAAAACGGTACTTATAACAATAGTTTGTGTTATGGGTTTATGTCCGGAATGGTAAT
AATAATCAAAATGAAATCAAGATGGCTTCTACATTGGGGATGTCTTTGGAAGGAACGAG
AGAAAGCAAGAGAATGTGAACAATAACAATAAATACTCAGAGAATCCTAGCAAGGTGTTT
TGGGGGTTTCCATGGCAGATGACCGGAGATTCCGCCGGAGTTGTACCGGAGATTGATCCC
GGAAGGGAAAGCTGGAATGGGATGCTTTCATCTTGAATAATGGTTTACTCAACACTCCT
TTGGTCTAGCAGATCATTAA
>G1886 Amino Acid Sequence (domain in aa coordinates: 17-59)
MDPEQEISNETLETILVSSTKGSNNNNKMEEMKKKVSRLGELGGEAQNCPRESPTNTKF
CYNNYSLSQPRYFCAACGCGGTATTTCTTGGCGGTGGAGATGGGCTTAGTCTAAGCCAG
GGCCACTTTCTTTTGTCTCTTCTTTGTTTGTGATAACTTCCCTGACGTCTTGACGCCGG
ATAACCAACGACGACGCTGTTTACTCAGCTTCTTAACGGAATATGTGGTGTCTCCTG
GTGGCGGAGGACGTTCAACGGCGGGGATGTTCCGCCGAGGAGTCCGATGTTTACAATCC
CTTCTGGTTTCAGCCCTTCTAGTCTTCTCACCTCGCCCATGTTCTTTCCCCCGCAGTCGT
CAGCTCATAACCGGCTTTATTCAACCACGGCAGCAGTCAACCGCAACCACAACGACCAG
ACACGTTTCTCACCATATGCCACCATCGACATCCGTCGCCGTCATGGTCGTCAATCTT
TAGACGTTTCAACAAGTAGATCAAGAGCTCGAAACCATTATAATAATCCGGGGAATAACA
ATAATAACCGGTGCTATAACGTTGTGAACGTTGATAAACCAGGCGGATGACGGTTATAACT
GGAGGAAGTACGGACAAAAGCCTATCAAAGGGTGTGAATATCCAAGGAGTTATTACAAAT
GTACACATGTTAACTGTCCGGTGAAGAAGAAAGTGAACGGTCATCGGATGGACAGATCA
CTCAGATCATTTACAAAGGTCAACATGATCACGAGAGGCCTCAGAATCGCCGTGGCGGTG
GAGGCAGAGATTCCACTGAGGTTGGTGGTGCAGGGCAAATGATGGAATCTAGTGATGATA
GTGGTTATCGTAAGGATCATGATGATGATGATGATGATGATGAAGATGATGAAGATCTTC
CGGCTTCAAAGATAAGAAGATAGACGGTGTGTCGACGACTCACCAGCGGTGACCGAGC
CTAAGATTATCGTTTCAGACAAAAGTGAAGTGCATCTTCTCGACGATGGCTATAGGTGGC
GTAAGTACGGACAAAAGTGTCAAAGGAAATCCCCATCCAAGGAGCTATTATAAATGTA
CAACGCCAAATTGTACGGTCCGTAAACATGTAGAGAGAGCTTCCACGGATGCTAAGGCTG
TGATTACAACCTTACGAAGGTAAACACAATCACGATGTCCCTGCCGCTAGAAACGGTACCG
CGGCAGCAACCGCAGCTGCGGTGGGGCCGTCTGACCACCATCGTATGAGATCAATGTCCG

GGAACAATATGCAACAACATATGAGTTTCGGTAACAATAAACACAGGCCAATCTCCGG
TTCTTTTGGAGTTGAAAGAAGAGAAAATCACAATTGACTTTTAAGAACCAAAGATTTCG
AGATTGATATT

>G1933 Amino Acid Sequence (conserved domain in AA coordinates:205-263, 344-404)

MAVEDDVSLIRTTTLVAPTRPTITVPHRPPAIETAAAYFFGGGDGLSLSPGPLSFVSSSLFV
DNFPDVLTPDNQRTTSFTQLLNGTMSVSPGGGGRSTAGMFAGGGPMFTIPSGFSPSSLLT
SPMFFPPQSSAHTGFIQPRQSQPQPQRPDTPPHMPPSTSVAVHGRQSLDVSQVDQRRAR
NHYNNPGNNNNNRSYNVNVNDKPADDGYNWRKYGQKPIKGCEYPRSYKCTHVNCVPVKKK
VERSSDGQITQIIYKGQHDHERPQNRGGGGRDSTEVEGGAGQMMESSDDSGYRKDHDDDD
DDDEDDDELPAKIRRIDGVSTTHRTVTEPKIIIVQTKSEVDLLDDGYRWRKYGQKVVKGN
PHPRSYKCTTPNCTVRKHVERASTDAKAVITTYEGKHNHDVPAARNGTAAATAAAVGPS
DHHRMRSMGNNMQHMSFGNNNNNTGQSPVLLRLKEEKITI*

>G2059 (58..1089)

TTAAGAACAGGCTTCATCTCTGACAAACACTCAAAAAACAAACAAAAAGGAACATG
GAAGATCAGTTTCTTAAATAGAACTAGCTTCATGCACGACAAGCTCTTGCTCTCTGGA
ATCTACGGGTTCTTGAGTTCTTCGACGCGCCCAACTTCTCGGTGTTCGAATATTTTGG
GAAGGTATGAAATCTCTCTCTCTCTGCTTCTTCGACTCCGAGCTACTTTGTGTCGCCT
CATGATCATGAGCTCACATCTTCTATTCATCCATCTCCGGTAGCTTCTGTTCCTTGAAC
TTTCTAGAATCTTTCTCTCAGTCTCAACATCTGATCATCATCTTCTAAACCTCCAAAC
CTTACTTTGTTCCTTAAAGAACCAAGCTACTAGAAGTTTCTCAATCCGAAAGCAACATG
AGCCCTTACCATAAAATACATCCAACTCTTTTATCAATCAGACCAAAACAGAAACGAA
TGGGTAGAGATCAATAAACTCTAACCAACTATCCCTCGAAAGTTTGGAAACTATTGG
CTAAGTACCACCAAGACTCAACCCATGAAGTCAAAAAACAAGAAAGGTTGTTTCAGACGACG
ACCCCAACAAACTGTATAGAGGAGTGAGACAAAGACACTGGGGCAAATGGGTGCGAGAG
ATTAGGCTTCCAAGGAACAGAACCCGTGTTTGGCTCGGCACCTTTGAAACCGCTGAGCAA
GCAGCAATGGCTTACGATACAGCAGCTTATATCTTCTGCGGAATTTCGCACACCTCAAC
TTTCTGATCTTAAACACCAGCTCAAGTCCGGTTCTTTGCGATGCATGATCGCCTCACTT
TTGGAGTCCAAGATTCAACAGATCTCATCTTCCCAAGTAAGTAAGTCTCTCTCTCTCTCT
CCTCCAAAGTGGGAACACCGGAGCAAAAGAATCATCATGAGATGGAGTCAGGAGAA
GACGTGATGATGAAGAAACAGAAAGCCATAAGGAAGTGATGGAAGGAGATGGTGTACAA
TTGAGTAGGATGCCTTCTTTGGATATGGATCTCATTGGGATGCTCTCTCATTTCTCTCAT
TCTTCTTGACTTCAAATTAATATTGTCAAACCTATTTTACTTACTTCTACCTTTTTTTA
TATCAAAAGTTTCCACCAAGAAAGAAATTCATATTATGATGCCAAGATTGGTTTGCATT
TGGGGTTGAACACATTGTAATCTTCTTACGACCACATAATCAAGTGGTCTCTCTTTTTT
TGCTCTGCTAA

>G2059 Amino Acid Sequence (conserved domain in AA coordinates:184-254)

MEDQFPKIETSMHDKLLSSGIYGLSSSTPPQLLGVPILFLEGMKSPLLPASSTPSYFVS
PHDHELTSSIIHSPVASVPWNFLBSPQSQPHDHPKPPNLTFLKEPKLLELSQSESN
MSPYHKYIPNSFYQSDQNRNEWVEINKLTINYPSKGFNYWLSSTTKTQPMKSKTRKVQT
TPTKLYRQVRQRHWGKWVAEIRLPRNRTRVWLGTFFETAEQAAMAYDTAAYILRGEFAHL
NFPDLKHQLKSGSLRCMIASLLESKIQQISSQVSNSPSPPPPKVGTPEQKNHMKMESG
EDVMMKKQKSHKEVMEGDGVQLSRMPSLDMDLIWDALSFPHSS*

>G2105 (42..1487)

CTCTCTGACTTGAACTCTTCTCTCTACCGAATCAAACCAATGGAGGATCATCAAACC
ATCCACAGTACGGTATAGAACACCATCTTCTCAATTCTCTCTGATCTCTTCGGCTTCA
ACCTCGTTTCAGCGCCGGACCAGCACCATCGTCTTCATTTCACCGACCATGAGATAAGTT
TATTGCCACGTGGAATACAAGGGCTTACGGTGGCTGGAAACAACAGTAACACTATTACAA
CGATCCAGAGTGGTGGCTGTGTGGTGGGTTTAGTGGCTTTACGGACGGCGGAGGAACAG
GGAGGTGGCCGAGGCAAGAGACGTTGATGTGTTGGAGGTCAGATCTCGTCTTGATCACA
AGTTCAAAGAAGCTAATCAAAGGGTCTCTCTGGGATGAAGTTTCTAGGATTATGTCGG
AGGAACATGGATACACTAGGAGTGGCAAGAAGTGTAGAGAGAAGTTCGAGAATCTCTACA
AGTACTATAAAAAACAAAAGAAGGCAATCCGGTCCGCGACAAGATGGTAAAACTATA
GATTTTTCCGGCAGCTTGAAGCGATATACGGCGAATCCAAAGACTCGGTTTCTTGCTATA
ACAACACGCAGTTCATAATGACCAATGCTCTTCATAGTAATTTCCGCGCTTCTAACATTC
ATAACATCGTCCCTCATCATCAGATCCCTTGATGACCAATACCAATACTCAAAGTCAAA
GCCTTAGCATTTCTAACAAATTTCAACTCTCTCTCGGATTTGGATCTAACTTCTCTCTG
AAGGAAACGAAACTACTAAAAGAGAGGGGATGCATTGGAAGGAAAAGATCAAGGAATTCA

TTGGTGTTCATATGGAGAGGTTGATAGAGAAGCAAGATTTTGGCTTGAGAAGTTGATGA
AGATTGTGGAAGACAAAGAACATCAAAGGATGCTGAGAGAAGGAATGGAGAAGGATTG
AAGCGGAAAGGATCGATAAGGAACGTTCTGTTTGGACAAAAGAGAGGGAGAGGATTGAAG
CTCGGGATGTTGCGGTGATTAATGCCTTGACGACTTGACGGGAAGGGCATTGATAAGGC
CGGATTCTTCGTCTCCTACAGAGAGGATTAATGGGAATGGAAGCGATAAAATGATGGCTG
ATAATGAATTTGCTGATGAAGGAAATAAGGGCAAGATGGATAAAAAACAAATGAATAAGA
AAAGGAAGGAGAAATGGTCAAGCCACGGAGGGAATCATCCAAGAACCAGAGAATATGA
TGATATACAACAATCAAGAACTAAGATTAATGATTTTGTGAGATGATGACCAATGCC
ATCATGAAGGTTACTCACCTTCAAACCTCCAAGAACGCAGGAACCTCCGAGCTGCAGCAATG
CCATGGCAGCTAGTACAAAGTGCCTTCCATTGCTTGAAGGAGAAGGAGATCAGAACTTGT
GGGAGGTTATGGTTTGAAGCAAAGGAAAGAAAATAATCATCAGTAAGCTACATTTTCA
TTCTCAAATGAAGAATAAGAGAACTTAGAAACGAT

>G2105 Amino Acid Sequence (domain in AA coordinates: 100-153)
MEDHQNHPOYGIEQSSQFSSDLFGFNLVSPDQHHRLHFTDHEISLLPRGIQGLTVAGN
NSNTITTIQSGGCVGGFSGFTDGGGTGRWRPQETLMLLEVRSLDHKFKEANQKGPLWDE
VSRIMSEEHGYTRSGKKCREKFENLYKYYKTKEGKSGRRQDGNRYRFRQLEAIYGESK
DSVSCYNNTFIMTNALHSNFRASNIHNIVPHHQNLMTNTNTQSQSLISNNFNSSDL
DLTSSSEGNETTKREGMHWKEKIKEFIGVHMERLIEKQDFWLEKLMKIVEDKEHQRLRE
EEWRRIEAERIDKERSFWTKERERIEARDVAVINALQYLTLGRALIRPDSSSPTERINGNG
SDKMMADNEFADEGNKGMDDKQMNKKRKEKWSHGNHPRTKENMMIYNNQETKINDFC
RDDDQCHHEGYSPSNSKNAGTPSCSNAMAASTKCFPLLEGGDQNLWEGYGLKQQRKENNH
Q*

>G2117 (49..465)
ATACTGTGCAACAAAAATTTCTTAAAGAACGCATAACTGTTTTTTTCATGGCTGGTTCT
GTCTATAACCTTCCAAGTCAAAACCCCTAATCCACAGTCTTTATTCCAAATCTTTGTTGAT
CGAGTACCACTTTCAAACCTGCTGCCACGTGACGACTCTAGCCGGACTGCAGAAGAT
AATGAGAGGAAGCGGAGAAGGAAGGTATCGAACCGCGAGTCAGTCCGGAGATCGCGTATG
CGGAAACAGCGTCACATCGAAGAACTGTGGTCCATGCTTGTTCAACTCATCAATAAGAAC
AAATCTCTAGTCGATGAGCTAAGCCAAGCCAGGGAATGTTACGAGAAGGTTATAGAAGAG
AACATGAACTTCGAGAGGAAAACTCCAAGTCGAGGAAGATGATTGGTGAGATCGGGCTT
AATAGGTTTCTTAGCGTAGAGGCCGATCAGATCTGGACCTTCTAATCGTCTCGTAAGCTT
GTTGGTTTTTTGTTGTTTATTAAAG

>G2117 Amino Acid Sequence (conserved domain in AA coordinates:46-106)
MAGSVYNLPSONPNPQSLQIFVDRVPLSNLPATSDSSRTAEDNERKRRRKVSNRESAR
RSRMKQRHMEELWSMLVQLINKNKS LVDLSQARECYEKVIEENMKLREENSKSRKMIG
EIGLNRFLSVEADQIWF*

>G2124 (87..923)
GAACAGCAAAAACCTAGATTTCCTGTTCAAGCTCAAGACCGTACAAAACCTTTGGAATCA
TATATAAAGATCTCGAGAATAGCATTATGAATATCGTCTCTTGGAAGATGCAAACGACG
AAGTTGCAGGCGGCGCTACGACAAGACGTGAAAGAGAAGTAAAAGAGGATCAAGAAGAAA
CCGAAGTCAGAGCCACCAGTGGCAAAACCGTAATTAAGAGCAGCCTACATCGATCTCTT
CTTCTTCTTCTTCTGTTGATGAAATCCAAGGATCCGAGGATTGTTAGGGTTTCACGCGCT
TTGGAGGCAAGACCGTCAAGCAAGTGTGTACGTTACGTGGACTACGTGACAGACGCG
TGAGATTATCAGTCCCAACGGCTATTACGCTCTACGATCTTCAAGAACGGCTCGGTGTG
ACCAGCTATGCAAAAGCCGTTGACTGGTTGCTTGATGCAGCTAAAGAGGAGATCGACGAGC
TACCTCCGTTACCTATCTCGCCGGAAAAATTCAGCATCTTCAACCATCATCAGTCTTCT
TGAATCTTGGTCAACGGCCCGGTCAAGATCCGACCAACTCGGGTTTAAAATCAATGGAT
GTGTACAAAAGTCTACTACTACTAGCCGCGAAGAAAACGATAGAGAGAAAGGAGAAAACG
ATGTCGTTTTACAAACAATCATCATGTTGGGTCTTATGGAACCTATCACAACCTGGAAC
ATCATCATCATCATCAACATTTGAGTTTACAGGCAGATTATCATAGTCATCAACTAC
ATAGTCTTGTCCCATTTCCATCACAATTTTGGTATGTCCAATGACGACATCAACCAAC
CTACAATATACAATCTTTGTTTCCATCATCATCGTCAGCTGGTTTCAGGGACTATGGAGA
CATTAGATCCGAGGCAATGTAGCAACAATGGTGGTAGAGACATTGATAATCGGATGTCG
TCGGTCCAATTAACCGAACTAATAGCACTACAACGGCTAACATGTCGAGGCATCTAGGC
TCGGAGCGTTGTACAAGTAGAGGAAGTGATCACCATATGTGAAGTTAGATTATTGAAACG
ATATAATTGTTGTTTGTGTTTCAGAAATAAGGGGACAC

>G2124 Amino Acid Sequence (domain in AA coordinates: 75-132)

MNIVSWKDANDEVAGGATRRREREVKEDQEETEVRATSGKTVIKKQPTSISSSSSSWMKS
KDPRIVRVSRAFGGKDRHSKVCTLRGLRDRRVRVLSVPTAIQLYDLQERLQVDQPSKAVDW
LLDAAKEEIDELPPLPI SPENFSIFNHHQSFLNLGQRPGQDPTQLGFKINGCVQKSTTTS
REENDREKGENDVVYTNHHVGSYGTYNLEHHHHHHQHLSLQADYHSHQLHSLVPFSPQ
ILVCPMTTSPTTTTIQSLFPSSSSAGSGTMTETLDPQM*

>G2140 (148..1254)

ACTCTCTTAACCTTTCTGTTCTTCTCTACCTTCTTTTACCAACCTTCTCTTCTTCTTACA
CACATATATATATACATATATAGAGAGAGAGAAGAGGACAAAGAGTTGAAAGATGAAGAC
TCTCATGTCTTCATAGAAACAAGTGATATGTGCGCTAAGAAAGAAGAAGAAGAAGAAGAA
GAAGAAGACAGTTCTGAAGCCATGAACAACATACAAAATTACCAAAATGACCTCTTCTTT
CACCAACTCATCTCTCATCATCACCATCATCATCATGATCCTTCTCAATCTGAAACTTTG
GGAGCATCCGGTAACGTTGGATCTGGTTTCACTATCTTCTCTCAAGATTCCGCTCTCTCCA
ATATGGTCTCTACCTCCACCTACCTCGATCCAACCACCATTTGATCAGTTTCTCTCTCTCT
TCTTCTTCTCCAGCATCTTTCTACGGAAGTTTCTTCAACAGAAGTCGAGCTCATCATCAG
GGATTACAGTTTGGGTACGAGGGTTTGGTGGAGCCAGTCAGCAGCACATCATCATCAT
GAACAACCTCGGATCTTGTGCGAAGCTTTAGGTCCGGTAGTACAAGCCGGGTCCGGTCTCT
TTTGGGTACAAGCTGAGTTAGGGAAGATGACAGCACAAGAGATCATGGACGCTAAAGCT
TTGGCTGCTTCAAAGAGTCATAGTGAAGCTGAGAGAAGAAGAAGAGAGAGAATCAATAAT
CATCTCGCTAAGCTCCGTAGCATATTACCCAACACCACCAAAACGGATAAAGCGTCGTTA
CTAGCTGAAGTGATCTCAACATGTGAAAGAGTTGAAGAGAGAGACTTCAGTGATCTCAGAG
ACAAATCTTGTCCTCAACGGAAGCGATGAGTTAACGGTAGCTTTCACGGAGGAGGAAGAA
ACCGGAGATGGCAGATTTGTAATTAAAGCGTCGCTTTGCTGTGAAGACAGGTCCGATCTC
TTGCCGTGACATGATTTAAACATTGAAAGCTATGCGTCTCAAACGCTCAAGGCGGAGATA
ACCACCGTTGGGGGACGAGTCAAGAACGTTTTTGTGTTACCGGAGAAGAGAGCTCCGGT
GAGGAAGTGGAGGAAGAGTACTGTATAGGGACGATTGAGGAAGCTTTGAAAGCGGTGATG
GAGAAGAGCAATGTAGAGGAATCATCTTCTTCTGGAAATGCTAAGAGACAGAGAATGAGT
AGTCACAACACTATCACTATCGTCGAACAACAACAATATAATCAGAGGTAATCAATT
TTTTACTTAAATCGCTTTTTTTTTCTTACTTTTCGGTGTATCTACTACGTGTGTTGTTGCT
GGTTATGGAATGAATGTTGTACGTCACGTTATACTATAGATATATGTGTGTTGTGTGT
ATGTATAACGGAAGTATTTGTATCCGTTGTGGTCTTGGACTTTTGGTTTGGTTCTAAGAT
ACTTATTTTTTAAAACTTGTATCGTTGAGTTGGTTTTCTAGATATGCTTAATGGGAGTAT
GTGACGAAAAAAA

>G2140 Amino Acid Sequence (domain in AA coordinates:167-242)
MCAKKEEEEEEDSSSEAMNNIQNYQNDLFFHQLISHHHHHHHHDPQSSETLGASGNVGS
FTIFSQDSVSPISLPPPTS IQPPFDQFPSPSSSPASFYGSFNRRAHHQGLQFGYEGF
GGATSAHHHHEQLRILSEALGPVVQAGSGPFLQALGKMTAQEIMDAKALAASKSHSE
AERRRRERINNHLAKLRSILPNTTKTDKASLLAEVIQHVKEKRETSVISETNLVPTESD
ELTVAFTEEEETGDRFVIKASLCCEDRSDLLPDMIKTLKAMRLKTLKABITTVGGRVKN
VLFVTGEESSEEBEYICGTIEEALKAVMEKS NVEESSSSGNAKRQRMSSHNTITIVE
QQQQYNQR*

>G2144 (102..1241)

ATTAGGGTTTTGTGTCGTGAGATTTGATTACACAAATTGCTGAATTTGGTTTTCGATTAT
TGGTGTATTGTTTTCGAAGATTTCCAGTGAGTTTCCGTTTATGGATCTGACTGGAGGAT
TTGGAGCTAGATCCGGCGGTGTTGGACCGTGCCGGGAACCAATAGGCCTTGAATCGCTAC
ATCTCGGTGACGAATTTCCGCAACTAGTGACGACTTTACCTCCCGAGAACCCCGCGGTT
CGTTCACGGCTTTGCTTGAGCTTCCACCTACACAAGCAGTGAGGCTTCTCCATTTCACTG
ATTCTTCGTCTTCTCAACAAGCGGAGTGACAGGGATCGGTGGAGAGATTCTCCGCCGC
TTCACTCTTTCCGGTGGGACATTTGGCTTTTCTTCTAAGTCACTGATTCATGGAGCGAGCAG
CTCGTTTCTCGGTGATTGCCACTGAGCAACAAAACGGAATATCTCCGGGAGACTCCGA
CGAGCTCTGTACCTTCAATCAAGTGCTAATCTCGACAGAGTCAAGACGGAGCCTGCTG
AGACCGATTCTCTCAGCGGTTGATTTCTGATTACGCGATTGAGAATCAAATCCCTTGCC
CTAACCAGAACAAATCGAAATGGGAAGAGGAAAGATTTGAAAAGAAGGGTAAAAGCTCGA
CGAAGAAGAACAAGCTCTGAAGAGAACGAGAAGCTGCCATATGTTACGTTAGAGCTC
GTCGTGGTCAAGCAACCGATAGCCATAGCTTAGCAGAACGAGCAAGAAGAGAGAAGATAA
ATGCACGAATGAAGCTGTTACAGGAACCTGGTCCAGGCTGTGATAAGATTCAAGGTACCG
CGCTGGTGTGATGAAATCATTAACCATGTCCAGTCATTACAACGTCAGTGGAGATGC
TATCAATGAGACTTGCTGCGGTAAACCCGAGAATCGACTTCAATCTCGACACCATATTGG

CTTCAGAAAACGGTTCCTTAAATGGATGGGAGCTTCAATGCCGCACCAATGCAGCTTGCTT
GGCCTCAGCAAGCCATTGAGACCGAACAGTCCTTTTCATCACCGGCACTGCAACAACCAC
CAACACAACAATGGCCTTTTGACGGCTTGAACCAGCCGGTATGGGGAAGAGAAGAGGATC
AAGCTCATGGCAATGATAACAGCAATTTGATGGCAGTTTCTGAAAATGTAATGGTGGCTT
CTGCTAATTTGCACCCAAATCAGGTCAAATGGAGCTGTAAGTTGGGAAAACGGTAGAGA
TCATGAATGTGTATATACATCGTATAAGCTCGTTTCTCTCTATATAAATATAATCATAAA
TATAGATATCTGTTAAGAAGGTATCAGTCATTTGATTTCAGAGAGACAACACTGGTATGAT
TGTTTCTTATTCTTGACCAGATTTGACAATGTAGAATTTAGTAGGATATGATCATTTT
GATCTCGTTATATATA

>G2144 Amino Acid Sequence (domain in AA coordinates:203-283)
MDLTGGFGARSGGVGPCREPIGLES LHLGDEFRLVTTLPENPGGSFTALLELPPTQAV
ELLHFTDSSSSQQAAVTGIGGEIPPLHSFGGTLAPPSNSVLMERARFSVIATEQQNGN
ISGETPTSSVPSNSSANLDRVKTEPAETDSSQRLISDSAIENQIPCPNQNNRNGKRKDFE
KKGKSSTKKNKSSEENEKLPYVHVRARRGQATDSHSLAERARREKINARMKLLQELVPGC
DKIQGTALVDEIINHVS LQRQVEMLSMRLA AVNPRIDFNLDTILASENGSLMDGSFNA
APMQLAWPQQA IETEQSFHHRQLQPPPTQQWPF DGLNQPVWGREEDQAHGNDNSNLMAVS
ENVMVASANLHPNQVKMEL*

>G2431 (47..1057)
CCCTTTTCGTTTATTATTTAAATTTCTTGGGTGCTTTC'TTAAATTTGTATGTGTTTATTAAT
GGAGATCAACAATAATGCCAACAATACTAATACTACTATTGATAATCACAAGGCAAAGAT
GAGCCTTGTTGTGTAACGGATGCTAAGCCAAGGTTGAAATGGACTTGATGATCTTCATCA
CAAATTCATCGAAGCCGTTAATCAACTTGAGGACCTAACAAAGCAACACCTAAGGGTTT
GATGAAGGTTATGGAGATTCTCGGCTTACCTTATACCATCTCAAGAGCCATTTACAGAA
ATATCGGTTAGGGAAGAGCATGAAGTTCGATGATAACAAGCTAGAAGTTTCTCTGCATC
AGAGAATCAAGAAGTTGAGAGTAAAAACGATTCAAGAGATCTCCGAGGCTGCAGTGTAC
CGAAGAAAACAGCAATCCAGCTAAAGAAGGGCTACAAATCACAGAGGCTTTACAAATGCA
GATGGAAGTTTCAAGAAGAACTTCATGAACAAATCGAAGTTCAGAGGCATTTGCAGGTGAA
GATTGAGGCACAAGGAAAGTATCTACAGTCCGTTTTAATGAAAGCTCAACAAACTCTCGC
TGGCTACTCATCTTCAAATCTCGGCATGGATTTTGCAGAGACCGAGCTCTCTAGATTAGC
TTCAATGGTTGAACGAGGCTGTCCAAGCACTTCGTTTCTCAGAGCTAACGCAAGTAGAAGA
AGAAGAAGAAGGTTTCTTGTGTTACAAGAAACCAGAAAACAGAGGAATTAGTCAGCTGAG
ATGTTTCAGTAGAGAGCTCGTTGACATCTTCAGAGACCTCAGAGACAAAACCTGGATACTGA
CAATAACCTTAATAAATCGATTGAACTTCCGTTGATGGAGATCAACTCGGAAGTGATGAA
GGGGAAGAAGAGAAGCATAAACGACGTCGTTTGCCTGGAGCAGCCTCTAATGAAGAGAGC
TTTTGGAGTTGATGATGATGAGCATTGGAAGTTGAGTTTGAATACTTACAAGAAAGACAT
GGAGGCGTGTACGAACATAGGACTAGGTTTAATTAATAAAAAAAAAACATTTTACTAAAGTT
ATATAAAAATGTTTTTAAAGAATCCA

>G2431 Amino Acid Sequence (conserved domain in AA coordinates:38-88)
MCLLMEINNNANNTNTTIDNHKAKMSLVLSTDAKPRLKWTCDLHHKFIEAVNQLGGPNKA
TPKGLMKVMEIPGLTLYHLKSHLQKYRLGKSMKFDDNKLEVS SASENQEVESKNDSRDLR
GCSVTEENS NPAKEGLQITEALQMOMEVQKKLHEQIEVQRHLQVKIEAQGKYLSVLMKA
QQTLAGYSSSNLGMDFARTELSRLASMVNRCGPSTSFSELTQVEEEEGFLWYKKPENRG
ISQLRCSVESLTSSETSETKLDTDNLNKSIELPLMEINSEVMKGKKRSINDVVCVEQP
LMKRAFGVDDDEHLKLSLNTYKKDMEACTNIGLGFN*

>G2465 (86..1150)
CAATATTCTTTCTCCATTGAGATTAAGCTTCTTTCTCGCTGTCGTCTCTCTATAGATCTT
GGTTCTTAGTCCCTTTTGAATAATAATGATGGTGGAGATGGATTACGCTAAGAAAATGCA
GAAATGTATGAATACGTTGAAGCACTTGAAGAAGAACAGAAGAAAATCCAAGTCTTTCA
ACGCGAGCTTCCCTTTATGTTTAGAGCTTGTCACCTCAAGCGATCGAAGCTTGTCGGAAGGA
GTTATCTGCTACGACGCAACTACATCAGAACAGTGTTTCAAGAACAGACCACAAGTGTGTTG
TGGTGGTCCCTGTCTTTGAAGAGTTTATTCTTATCAAGAAAATTAGTTCTTGTGTGAAGA
AGTACAAGAAGAAGAAGAAGATGGTGAACATGAATCTTCTCCAGAACTTGTAATAA
TAAGAAATCAGATTGGCTTAGATCTGTTTCAAGCTATGGAATCATTACCCGGATCTAAATCC
AAAAGAGGAGCGTGTAGCTAAGAAAGCGAAAGTGGTGGAGGTGAAACCAAAAAGCGGTGC
GTTTCAGCCGTTTCAAAAGCGCGTTTGGAGACTGATTTGCAACCGCGGTGAAAAGTAGC
TAGTTTCGATGCCAGCGACGACGACGAGTTCTACGACGGAACTTGTGGTGGTAAAAGTGA
TTTGATTAAAGCTGGAGATGAGGAAAGACGGATAGAGCAGCAGCAATCGCAGTCGCATAC

GCATAGAAAACAAAGGCGGTGCTGGTCGCCGAATTACACCGTCGATTCTTAAACGCGCT
TCAGCAGCTTGGAGGATCTCATGTTGCTACACCAAAGCAAATCAGGGATCACATGAAGGT
TGATGGATTAAACAAACGACGAAGTTAAAGCCATTTACAGAAATATAGACTTCACACAAG
AAGGCCAGCAGCAACATCCGTGGCGGCACAAAGTACCGGAATCAGCAACAACCACAATT
TGTGGTGGTTGGAGGCATATGGGTACCATCGTCACAAGATTTCCACCACCGTCCGATGT
AGCCAACAAGGGTGGTGTATATGCTCCGGTTGCGGTGGCGCAATCTCCAAAACGTTCTGTT
GGAGAGAAGTTGCAACTCGCCGGCGGCATCTTCTCTACAAATACAAATACTTCTACTCC
TGTGTCATAATCTGATAGTCATACTATAATCATCTCTGATGTTGATTTTGGTGTAGGTT
TGAAAATGTTTATGTGAATGTAA

>G2465 Amino Acid Sequence (conserved domain in AA coordinates:219-269)

MMVEMDYAKMKQKCHEYVEALEEEQKKIQVFQRELPLCLELVTQAI EACRKELSGTTTTT
SEQCSEQTTSVCGGPVFEEFIPIKKISSLCEEVQEEEEEDGEHESPVLVNNKSDWLRS
VQLWNHSPDLNPKEERVAKKAKVVEVKPKSGAFQPFQKRVLETDLQPAVKVASSMPATTT
SSTTETCGGKSDLIKAGDEERRIEQQQSQSHTRKQRRCSPELHRRFLNALQQLGGSHV
ATPKQIRDHMKVDGLTNDEVKSHLQKYLHTRRPAATSVAQSTGNQQQPQFVVVGGIWW
PSSQDFPPPSDVANKGGVYAPVAVAQSPKRSLSRSCNSPAASSSTNTNTSTPVS*

>G2583 (38..607)

CAAAATCAGAAAATATAGAGTTTGAAGGAACTAAAAGATGGTACATTGAGGAAGTTCCG
AGGTGTCCGCCAGCGACATGGGGTTCTTGGGTCTCTGAGATTGCGCATCTCTTATTGAA
GAGAAGAGTGTGGCTTGGAACTTTGGAACCGGCAGAAGCGGCTGCAAGAGCATACGACCA
AGCGGCTCTTCTAATGAACGGCCAAAACGCTAAGACCAATTTCCCTGTCTGTAATAATCAGA
GGAAGGCTCCGATCACGTTAAAGATGTTAACTCTCCGTTGATGTCACCAAAGTCATTATC
TGAGCTTTTGAACGCTAAGCTAAGGAAGAGCTGCAAAGACCTAACGCCTTCTTTGACGTG
TCTCCGTCTTGATACTGACAGTTCCACATTTGGAGTTTGGCAGAAACGGGCGGGTTCGAA
AACAAGTCCGACTTGGGTGATGCGCCTCGAAGTTGGGAACGTAGTCAACGAAAGTCCGGT
TGACTTAGGGTTGACTACGATGAACAAACAAAACGTTGAGAAAGAAGAAGAAGAAGA
AGCTATTATTAGTGATGAGGATCAGTTAGCTATGGAGATGATCGAGGAGTTGCTGAATTG
GAGTTGACTTTTGACTTTAACTTGTGCAAGTCCACAAGGGGTAAAGGGTTTTTC

>G2583 Amino Acid Sequence (domain in AA coordinates:4-71)

MVHSRKFRGVRQRQWGSWSEIRHPLLRVWLGTFTAEAAARAYDQAALLMNGQNAKT
NFPVVKSEEGSDHVKDVNSPLMSPKSLSELLNAKLKSKDLTPSLTCLRLDTS SHIGV
WQKRAGSKTSPTWVMRLELGNVVNESAVDLGLTTMNKQNVKEEEEEEEAII SDEDLAME
MIEELLNWS*

>G2724 (1..651)

ATGGAAATAGAAATAAGGAGAGGTCCATGGACTGTGGAAGAAGACATGAAGCTCGTCAGT
TACATTTCTCTTACCGTGAAGGAAGATGGAACTCCCTCTCTCGTTCTGCTGGACTGAAT
AGAACGGGGGAAAAGTTGCAGATTGCGGTGGCTAAATTATCTCCGGCCGGATATCCGCCGT
GGAGACATATCCCTTCAAGAACAAATTTATCATCCTTGAAGTCCATTCTCGTTGGGGAAAT
CGGTGGTCAAAGATTGCTCAACATTTACCGGAAGAAGACAGATAACGAGATAAAGAATTAT
TGGAGAACACGTGTTCAAAGCATGCAAACTTCTAAAATGTGACGTGAACAGCAAGCAA
TTCAAAGACACCATCAAACATCTCTGGATGCCTCGTCTCATCGAGAGAATCGCCGCCACT
CAAAGTGTCAAATTTACCTCTAACCCTACTCGCCTGAGAACTCCAGCGTCGCCACCGCC
ACGTCATCAACGTCGTCGTCTGAGGCTGTGAGATCGAGTTTCTACGGTGGTGATCAGGTG
GAATTTGGAACGTTGGATCATATGACAAATGGTGGTTATTGGTTCAACGGCGGAGATACG
TTTGAACTTTGTGTAGTTTGTGACGAGCTCAACAAGTGGCTCATACAGTAG

>G2724 Amino Acid Sequence (conserved domain in AA coordinates:7-113)

MEIEIRRGPTVEEDMKLVSYISLHGEGRWNSLSRSAGLNRTGKSCRLRWLNLYLRPDIRR
GDISLQEQFIILELHSRWGNRWSKIAQHLPGRDNEIKNYWRTRVQKHAKLLKCDVNSKQ
FKDTIKHLWMPRLIERIAATQSVQFTSNHYPENSSVATATSSSSSEAVRSSFYGGDQV
EFGTLDHMTNGGYWFGNGDTFTFELCSFDELNKWLIQ*

>G377 (1..396)

atgggtctctcgcatTTTTCCAACAGCGTCAGAAGGAGTACTACCATTCTGGTGATGAAC
acgggtgtttcaatcactctgttgaagaacatggtagggtctgtttttcaaatgtttgca
tccgagactgaatcttccatggagatagacgacgagcctgaagatgattttgttactaga
agaatctcgataacacagttcaagtctctatgtgagaacatagaagaggaagaagaagag
aaaggtgtggaggtgtgtgtgtcctttgtgggttaaagaggaagaggaagtgagtgag
ttgggtttcttgcaagcattttctccacagagcttgtctagacaactgggtttgtaataac

cacaccacatgccctctttgcagggtccattctctag
>G377 Amino Acid Sequence (domain in AA coordinates:85-128)
MGLSHFPTASEGVLP LLVMNTVVSITLLKNMVRVVFQIVASETESSMEIDDEPEDDFVTR
RISITQFKSLCENIEEEEEKGVCCVCLCGFKEEEVSELVSCKHFFHRACLDNWFGNN
HTTCPLCRSIL*
>G428 (97..1032)
TTACTTTTGTGTTTCTTCATATTCTTCAGAAGCAAGCACAAAGGCTAGGGATCGAAGAAGC
GGCGATCACTGATCGTATCTCACTACGATCACATTAATGGATAGAATGTGTGGTTTCCGC
TCGACGGAAGACTATTTCGGAGAAAGCGACGTTGATGATGCCGTCCGATTATCAGTCTTTG
ATTTGTTCAACCACCGGAGACAATCAAAGACTGTTTGGATCCGACGAACCTCGTACCGCT
TTGTCCTCGGAGTTGCTTCCGCGTATTTCGAAAAGCTGAGGATAATTTCTCTCTTAGTGTC
ATCAAATCCAAATCGTTCTCATCTTTGTATCCTCGCTTACTCCAAACCTACATCGAT
TGCCAAAAGGTGGGAGCGCCTATGGAAATAGCGTGTATATTGGAAGAGATTGAGCGAGAG
AACCATGTGTACAAGAGAGATGTTGCTCCATTATCTTGCTTTGGAGCTGATCCTGAGCTT
GATGAATTCATGGAACCTACTGTGATATATTGGTTAAATACAAACCGATCTTGCAGAGG
CCGTTTCGACGAGGCTACAACCTTTTATAAACAAGATTGAAATGCAGCTTCAGAACTTGTGC
ACTGGTCCAGCGTCTGTACAGCTCTTTCAGATGATGGTGCGGTTTCATCTGACGAGGAA
CTGAGAGAAGATGATGACATAGCAGCGGATGACAGCCAAACAAAGAAGCAATGACCGCGAT
CTGAAGGACCAGCTACTACGCAAAATTTGGTAGCCATATCAGTTTCATTGAAACTCGAGTTC
TCTAAAAGAAGAAGAAAGGGAAGCTACCAAGAGAAGCAAGACAAGCGTTGCTCGATTGG
TGGAAATGTTTATAATAATGGCCTTACCCTACTGAAGGCGACAAATAGCTCTGGCTGAA
GAAACAGGTTTGGATCAAAAACAATCAACAATTGGTTTATAAACCAAGGAAACGCCAT
TGGAAAGCCTTCGGAGAACATGCCGTTTGTATGATGGACGATTCTAATGAAACATTCTTT
ACCGAGGAATGAAAAGAGAGACATGGGATTGTGCATTGTATAATTTTACACTGTTTTCC
CAAGAAAAGAAAACAGTAAAAGCTTTTGGTAAATGGGACATCATCGGAATGAATGGAA
CCAGTTAGCCAAAACGGTCAAGGGCGTGGCGTAACGAGACATTGTATTGAAATAGTGGC
AATATTATGTCTCAATCTTCCAATGGTCCAAAATGATAGATTTCTTATTGTATTGAAC
CTTACTTAGATAGCTGATGTGTCAACTAAATAATTTATTTTCATCCTTATACTACTTGTA
TCAATGTCTCTAATTGATCAATTGTTGCTTCTATTCAAAAAAAAAAAAAAAAAAAAAA
>G428 Amino Acid Sequence (domain in AA coordinates: 229-292)
MDRMCGFRSTEDYSEKATLMMPDSYQSLICSTTGDNQLRFGSDELATALSSELLPRIRKA
EDNFSLSVIKSKIASHPLYPRLLQTYIDCQKVGAPMEIACILEEIQRENHVKRDVAPLS
CFGADPELDEFMETYCDILVKYKTDLARPFD EATTFINKIEMQLQNLCTGPASATALSDD
GAVSSDEELREDDIAADDSQQRSDNRDLKDQLLRKFGSHISSLKLEFSKKKKKGKLPRE
ARQALLDWWNVHNKWPYPTEGDKIALAETGLDQKQINNWF INQRKRHWKPS ENMPFDM
DDSNETFFTEE*
>G447 (241..3501)
CTTTTAAAGAGCTTAAAAATTTGCTTTGAAGCTTCAAATATTCTTATGAACATAAAAGAA
GAAAAAGCTTTTGTCTTTTTCCTTAGCAGCAGAATGATTTTGTCTTCCAAAATTATT
ACTATTTAGTTTCTCTCGTGCTCTTCTCTTGAGCAAATACAGATTGCTTAATTTTGCTGA
AGAAGAAGAACTCTGTTTCTTCCCTGCACCAAACCAATTTTTCGTTCTTTCTATAAACC
ATGAAAGCTCCATCAAATGGATTTCTTCCAAGTTCCAACGAAGGAGAGAAGAAGCCAATC
AATTCTCAACTATGGCAGCTTGTGCAGGGCCTTTAGTTTCATTACCTCCTGTGGGAAGT
CTTGTGTTTACTTCCCTCAAGGACACAGCGAGCAAGTTGCAGCATCGATGCAGAAGCAA
ACAGATTTTATACCAAATTACCCAAATCTTCTTCTAAGCTGATTGCTTGTCTTACAGT
GTTACATTACATGCTGATACCGAAACAGATGAAGTCTATGCACAAATGACTCTTCAACCT
GTGAATAAGTATGATAGAGAAGCATTTGCTAGCTTCTGATATGGGCTTGAAGCTAAACAGA
CAACCTACTGAGTTTTTTTGAAGACTCTTACTGCAAGTGACACAAGCACTCATGGTGGA
TTCTCTGTACCGCGTCGTGCAGCTGAGAAAATATTCCTCCTCTTGATTTCTCGATGCAA
CCGCTGCGCAAGAGATTGTAGCTAAAGATTTACATGATACTACATGGACTTTTACAGCAT
ATCTATCGAGGCCAACCAAAAAGACACTTGCTTACCACAGGTTGGAGCGTTTTTGTGTTAGC
ACAAAGAGACTATTTGCGGGTGATTGAGTTTTGTTTGTAAAGAGATGAGAAATCACAGCTG
ATGTTGGGTATAAGACGTGCAATAGACAAACTCCGACTCTTTCCTCATCGGTATATCC
AGCGACAGTATGCACATTGGGATACTTGCAGCTGCAGCTCATGCTAATGCCAATAGTAGC
CCTTTTACCATCTTCTTCAATCCAAGGGCAAGTCCTTCAGAGTTTGTAGTTCTTTTAGCC
AAATACAACAAAGCCTTATACGCTCAAGTATCTCTAGGAATGAGATTCCGGATGATGTTT
GAGACTGAGGATTGTGGGGTTCGTAGATATATGGGTACAGTCACAGGTATTAGTGATCTT

GACCTGTAAAGATGGAAAGGCTCACAATGGCGTAATCTTCAGGTAGGATGGGATGAATCA
 ACAGCTGGAGATAGGCCAAGCCGAGTATCCATATGGGAAATCGAACCCTCATAACTCCT
 TTTTACATATGTCCTCTCCATTTTTCAGACCTAAGTACCCGAGGCAACCCGGGATGCCA
 GATGATGAGTTAGACATGGAAAATGCTTTCAAAAGAGCAATGCCTTGGATGGGAGAAGAC
 TTTGGGATGAAGGACGCACAGAGTTCGATGTTCCCTGGTTTAACTCTAGTTCAATGGATG
 AGTATGCAGCAAAACAATCCATTGTGTCAGGTTCTGCTACTCCTCAGCTCCCGTCCGCGCTC
 TCATCTTTTAACCTACCAAACAATTTTGCTTCCAACGACCCCTTCCAAGCTGTTGAACTTC
 CAATCCCCAAACCTCTCTTCGCAAAATTCCTCAATTCACAAACCGAACACGGTTAACCAT
 ATCAGCCCAACAGATGCAAGCACAACCAGCCATGGTGAAATCTCAACAACAACAACAACA
 CAACAACAACAACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA
 TCACAGCAACAGGTGCAGCAACAAGGGATTTATAACAATGGTACGATTGCTGTTGCTAAC
 CAAGTCTCTTGTCAAAGTCCAAACCAACCTACTGGATTCTCTAGTCTCAGTCTCAGCAG
 CAGTCAATGCTCCCTACTGGTGCTAAAATGACACACCAGAACAATAAATCTATGGGGAAT
 AAAGGCTTGTCTCAAATGACATCGTTTGCGCAAGAAATGCAGTTTCAGCAGCAACTGGAA
 ATGCATAACAGTAGTACCTATTATTAAGAAACAGCAAGAAGTCTCTCTCCATTCTATTA
 CAACAAAATCTGTCCCAAAATCCTCAGCAACTCCAAATGCAACAACAATCATCAAAACCA
 AGTCTTTCAACAACAGCTTCAGTTGCAGCTACTGCAGAAGCTACAGCAGCAGCAACAGCAG
 CAGTCGATTCTCCAGTAAGCTCATCCTTACAGCCACAATTATCAGCGTTGCAGCAGACA
 CAAAGCCATCAATTGCAACAACCTTCTGTGCTCTCAAAATCAACAGCCCTTGGCACATGGT
 AATAACAGCTTCCCAGCTTCAACTTTTCATGCAGCCTCCACAGATTTCAGGTGAGTCCCTCAG
 CAGCAAGGACAGATGAGTAACAAAAATCTGTAGCCGCTGGAAGATCACATTCTGGCCAC
 ACAGATGGAGAAGCTCTTCTTGTTCAAACCTCACCTTCCGCCAATAACACGGGACATGAT
 AATGTTTCAACGACAAATTTCTGTAGCAGAAATCAACAGCAAGGACAAGCTGCATCTGTA
 TCTGCATCTGATTAGTCTTTGAGCGCGCAAGCAATCCGGTCCAAGAGCTTTATACAAAA
 ACTGAGAGCCGGATCAGTCAAGGCATGATGAATATGAAGAGTGTGGTGAACATTTCAGA
 TTTAAAAGCGCGGTAAACAGATCAAAATCGATGTATCCACAGCGGAACGACGTAAGTCTCCT
 GATGTTGTTGGCCCTGTACAGCAGCAACAACTTTCCCACTACCATCATTTTGGTTTTGAT
 GGAGACTGCCAATCTCATCATCAAGAAACAACCTTAGCTTTCCCTGGTAATCTCGAAGCC
 GTAACCTTCTGATCCACTCTTATTCTCAAAAGGACTTTCAAAACCTTGGTTCCCAACTATGGC
 AACACACCAAGAGACATTGAGACGGAGCTGTCCAGTGTGCAATCAGTTCTCAGTCATTT
 GGTATTCCCAGCATTTCCCTTTAAGCCCGGATGTTCAAATGAGGTTGGCGGCATCAATGAT
 TCAGGAATCATGAATGTTGGAGGACTGTGGCCCAATCAGACTCAACGAATGCGAACATAT
 ACAAAGGTTCAAAAACGAGGGTCAGTAGGTAGATCAATAGATGTTACCCGTTATAGCGGC
 TATGATGAACCTTAGGCATGACTTAGCGAGAATGTTTGGCATCGAAGGACAGCTCGAAGAT
 CCGCTAACCTCTGATTGGAAACTCGTCTACACCGATCAGCAAAACGATATTTTACTAGTT
 GGTGATGATCCTTGGGAAGAGTTTGTGAAGTGCCTGCAGAACATAAAGATACTATCATCA
 GTAGAAGTTAGCAAAATGAGCTTAGACGGAGATCTTGCAGCTATCCCAACCAACAACCAA
 GCCTGCAGCGAAACAGACAGCGGAAATGCTTGGAAAGTACACTATGAAGACACTTCTGCT
 GCAGCTTCTTTCAACAGATAGAAATAAAAAGATGCAAAATATACCAAGTCAACTTACATTA
 TCATTCCAGGCCATCGCAAAGTACATGTTTTTTTTTGTGTGTATGTACTGCAAAACAACA
 ACTGAGAAGAAGAAGATACTGCACGGTATATAAACATTTTTTATAGGACAGTGATTGATT
 TTTTCAATTCTAACTTGATGTTGTTGTACTTTCTTGTTCATATTTGTATAACAAGTATAA
 TGCTTGACAAGTCTATGAGGAGCATATCTTATACAGAGATACTAAGATGTAATGTTAATG
 TAACTAAACAATTACCTTCATTAATCATGAATCCTTTGGTCGTTTAAAA

>G447 Amino Acid Sequence (conserved domain in AA coordinates:22-356)

MKAPSNGLPSSNEGEKKPINSQWLHACAGPLVSLPPVGSLLVYFPQGHSEQVAASMQKQ
 TDFIPNYPNLPKSLICLLHVSLLHADTETDEVYAQMTLQPVNKYDREALLASDMGLKLN
 QPTEFFCKTLTASDSTHGGFSVPRRAAEKIFPPLDFSMQPPAQEIVAKDLHDTTWTFRH
 IYRGQPKRHLLTTGWSVFVSTKRLFAGDSVLFVRDEKSQLMLGIRNRANRQPTLSSSVIS
 SDSMHIGILAAAAHANANSSPFTIFNPRASPEFVVPPLAKYNKALYAQVSLGMRFRMMF
 ETEDCGVRRYMGTVTGISDLDPVRWKGSRWNLQVWDESTAGDRPSRVSIWEIEPVITP
 FYICPPFFFRPKYPRQPGMPDDELDMENAFKRAMPWMGEDFGMKDAQSSMPGLSLVQWM
 SMQQNNPLSGSATPQLPSALSSFNLPNNFASNDPSKLLNFQSPNLSSANSQFNKPNTVNH
 ISQQMQAQAPAMVKSQQQQQQQQQHQHQQQQLQQQQQLQMSQQQVQQQGIYNNGTIAVAN
 QVSCQSPNQPTGFSQSQLQQQSMPLTGAKMTHQNINSMGNKGLSQMTSFAQEMQFQQQLE
 MHNSSQLLRNQEQSLHSLQNLQNPQQLQMQQSSKPSPSQQLQLQLQLQLQQQQQ
 QSIIPVSSSLQPLSALQQTQSHQLQQLLSQNPQPLAHGNNSFPASTFMQPPQIQVSPQ

QQGQMSNKNLVAAGRSHSGHTDGEAPSCSTSPSANNTGHDNVSPTNFLSRNQOQQAASV
SASDSVFERASNPVQELYTKTESRISQGMNMKSAGEHFRKSAVTDQIDVSTAGTTYCP
DVVGVPVQQQQTFFPLPSFGFDGDCQSHHPRNNLAFPGNLEAVTSDPLYSQKDFQNLVPNY
NTPRDIETELSSAAISSQSFQIPSIPIFKPGCSNEVGGINDSGIMNGGGLWPNQTRMRTY
TKVQKRGSVGRSIDVTRYSGYDELRHDLARMFGIEGQLEDPLTSDWKLVYTDHENDILLV
GDDPWEEFVNCVQNIKILSSVEVQQMSLDGDLAAIPTTNQACSETDSGNAWKVHYEDTSA
AASFNR*

>G464 (41..760)

CTCTGCTGGTATCATTGGAGTCTAGGGTTTTGTTATTGACATGCGTGGTGTGTCAGAATT
GGAGGTGGGAAGAGTAATCTTCCGGCGGAGAGTGAGCTGGAATTGGGATTAGGGCTCAG
CCTCGGTGGTGGCGCGTGGAAAGAGCGTGGGAGGATTCTTACTGCTAAGGATTTCTCTTC
CGTTGGGTCTAAACGCTCTGCTGAATCTTCTCTCACCAAGGAGCTTCTCCTCCTCGTTC
AAGTCAAGTGGTAGGATGGCCACCAATTGGGTTACACAGGATGAACAGTTTGGTTAATAA
CCAAGCTATGAAGGCAGCAAGAGCGGAAGAAGGAGACGGGGAGAAGAAAGTTGTGAAGAA
TGATGAGCTCAAAGATGTGTCAATGAAGGTGAATCCGAAAGTTCAGGGCTTAGGGTTTGT
TAAGGTGAATATGGATGGAGTTGGTATAGGCAGAAAAGTGGATATGAGAGCTCATTCGTC
TTACGAAAACCTTGGCTCAGACGCTTGAGGAAATGTTCTTTGGAATGACAGGTACTACTTG
TCGAGAAAAGGTTAAACCTTTAAAGGCTTTTAGATGGATCATCAGACTTTGTACTCACTTA
TGAAGATAAGGAAGGGGATTGGATGCTTGTGGAGATGTTCCATGGAGAATGTTTATCAA
CTCGGTGAAAAGGCTTCGGATCATGGGAACCTCAGAAGCTAGTGGACTAGCTCCAAGACG
TCAAGAGCAGAAGGATAGACAAAGAAACAACCCTGTTTAGCTTCCCTTCAAAGCTGGCA
TTGTTTATGATTGTTTGGAGTTTGAATTTACTCGATACTTTTGAAGAAAGTATTTTG
GAGAATATGGATAAAGCATGCAGAAGCTTAGATATGATTGAATCCGGTTTTTCGGATAT
GGTTTTGCTTAGGTCAATTCGTAGTTTTCAGTTTGTTCCTTTGCTTGGCTGTGTAC
CAATTATCTATGTTCTGTGAGAGAAAGCTCTT

>G464 Amino Acid Sequence (domain in AA coordinates: 20-28, 71-82, 126-142, 187-224)

MRGVSELEVGSNLPAESELELGLGLSLGGGAWKERGRILTAKDFPSVGSKRSAESSSHQ
GASPPRSSQVVGWPPIGLHRMNSLVNNQAMKAARAEEDGGEKKVVKNDKDVSMKVNPK
VQGLGFVKVNMMDGVGIGRKMRAHSSYENLAQTLEEMFFGMTGTTCREKVKPLRLLDGS
SDFVLTYEDKEGDWMLVGDVPMRMFINSVKRLRIMGTSEASGLAPRQEQKDRQRNNPV*

>G557 (192..698)

CAGAGATCTGACGGCGGTAGCAGAGTAATCTATTCCTTCCCAAATGTCTCGCAATTAGA
TTCTTTCCAAGTTCTTCTGTAAATCCCAAGTCCCGCTCTTTTCTCTTTATCCTTTTCAC
CAGCTTCGCTACTAAGACAACAAATCTTTCCCTCTCTCTCGCCTGATCGATCTTCAAA
GAGTAAGAAAAATGCAGGAACAAGCGACTAGCTCTTTAGCTGCAAGCTCTTTACCATCAA
GCAGCGAGAGGTCATCAAGCTCTGCTCCACATTTGGAGATCAAAGAAGGAATTGAAAGCG
ATGAGGAGATACGGCGAGTGCCGGAGTTTGGAGGAGAAGCTGTGCGAAAAGAACTTCGG
GTAGAGAATCTGGATCGGCGACCGGTGAGGAGCGGACACAGGCGACTGTGCGAGAAAGTC
AAAGGAAGCGAGGGAGGACACCGCGGAGAAAGAGAACAAGCGGCTGAAGAGGTTGTTGA
GGAACAGAGTTTTCAGCTCAGCAAGCAAGAGAGAGGAAAAAGGCTTACTTGAGCGAGTTGG
AAAACAGAGTGAAAGACTTGGAGAACAAAACTCTGAACCTGAAGAGCGACTCTCTACTC
TTCAGAACGAGAACAGATGCTTAGACATATTCTGAAGAACACAACAGGAAACAAGAGAG
GAGGTGGTGGTGGTTCTAATGCTGATGCAAGCCTTTGATCTCCTTCTTCTTCTTGTGTTA
TATTTTGTGGATAAAATTTACAGAGAATTGTATCAATAATTATCATGTTAAATTTATAT
GGGATGTGAGAGCTAATATTGCAATTGTAGACCAAGTTCTTTAAAAA

AAAAA

>G557 Amino Acid Sequence (domain in AA coordinates: 90-150)
MQEQATSSLAASSLPSSSSERSSSSAPHLEIKEGIESDEEIRRVPEFGGEAVGKETSGRES
GSATGQERTQATVGESQKRGRTPAEKENKRLKRLLRNRVSAQQARERKKAYLSELENRV
KDLENKNSLEERLSTLQENQMLRHILKNTTGNKRGGGGGSNADASL*

>G577 (44..2155)

AAAAACAGACTGAGAGAGAGAGAGAGAGTGTGTTGTTGGCCATGGGATGCACGGCCTC
CAAGCTCGACAGTGAGGATGCTGTCGCTGCAAGGAGCGGCGCGCTCTTATGAAGGA
CGCCGTCTACGCTCGTCACCATCTCGCCGCGCTCACTCTGACTACTGCGCTCCCTTCG
TCTACTGGCTCTGCCCTCTCCTCTTCGCGCGCGGCGAGCCCCTCTCCGTCTCCGAGAA
TACTCCGCTGTTTTCTCGCCCTTCTCCAGTCAGGACGCGCCACGTGTCCCTTCTTC

CCATTCCCCAGAACCCCTCCTCCGCCATCCGACGCAAGCCTAAGCCTACTAGGCCTAG
GAGGCTTCCACACATTCTCTCCGACTCCTCTCCTTCTTCTCTCTGACCAGTTTCTA
TCCCCTGCTCACCAGAACTCTACTTACTCTCGCTCTCCATCTCAAGCTTCTCTGTCTG
GAACTGGGAGAATTTCTACCTCCTCTCCCCCGACTCCGAGTACTTCGAACGCAAAGC
TCGCCAGAACACAGCACCGTCTCTCTCCGACTACGACGCCGAAACTGAAAGATCCGA
CCACGATTACTGCCACTCACGGAGAGATGCCGCCGAGGAAGTTCACTGCAGCGAGTGGGG
CGACGACCACGACCGTTTCTACTGCCACCTCTTCGTCCGACGGAGATGGGGAGGTGAAAC
TCACGTTTCCAGATCCGGTATTGAAGAAGAGCCTGTGAAACAACACATCAAGACCCAAA
TGGCAAAGAGCAGCTCTGACCATGTACACTTCTTCCGACTGCTACAAGACCAAATTGGT
GGTAAGGCACAAGAATTTGAAGGAGATCCTTGACGCCGTTCAAGACTACTTCGACAAGGC
TGCCTCCGCTGGGGACAGGTCTCCGCCATGCTTGAGATCGGCCGGGCTGAGCTCGACCG
CAGCTTCAGCAAGCTGAGGAAGACGGTGTATCATCAAGCAGTGTGTTTCAAGCACTTGAG
CGCAAGCTGGACCTCAAAACCCCATTTGGCAGTCAAATACAAGCTCGATGCATCTACCTT
GAATGATGAACAAGCGGCCCTCAAGAGCCTCTGCTCCACTCTAGACCGACTCCTCGCTTG
GGAAAAGAAGCTTTATGAGGATGTCAAGGCAAGAGAAGGAGTTAAGATTGAGCAGGAGAA
GAAGCTGTCTGCGCTGCGCTGAGCAGTATAGAGGAGTATAAGGGAGGTGATGAATCCAAGCTAGACAA
GACTAAAACCTTCCATAACAGACTGCAATCACTCATCATTGTTTCTTTCAGAAGCTGTTTT
AACCACGTCTAATGCCATTCTCCGCCCTCCGGGACACTGACCTTGTCCTCAGCTTGTTGA
ACTCTGCCACGGATTATGTACATGTGGAAGTCAATGCACGAGTATCACGAAATCCAGAA
CAACATCGTGCAACAAGTCCGTGGCCTGATCAACCAAACAGAGAGAGGTGAGTCAACATC
AGAGGTACACCGGCAGGTGACCGCGGACCTAGAGTCAGCTGTGTCTTGTGGCATTTCGAG
CTTCTGTGCGCATCATTAAATTCAGAGGGAGTTTATGCTCTCTCCACGCATGGTTCAA
GCTGAGCCTGGTTCCCTGAGCAACGGAGACCCAAAGAAACAGCGGCCAGACTCATTGTC
CTTGTGCGAGGAGTGAAGCAGAGCCTGGAACGGGTGCCTGACACAGTGGCGTCAGAAGC
CATAAAGAGCTTTGTAAACGTGGTACATGTGATATCAATAAAGCAGGCGGAAGAGGTGAA
GATGAAGAAACGCACGGAGAGTGCAGGAAAGGAGCTGGAGAAGAAAGCATCTCACTGAG
GAGCATAGAGAGGAAGTACTACCAGGCATACTCGACGGTTGGGATAGGCCCTGGACCGGA
GGTGTGGACTCACGGGACCCGCTATCTGAGAAGAAATGTGAGCTGGCGGCATGTGAGAG
GCAGGTGGAGGATGAGGTAATGAGGCACGTGAAGGCTGTGGAGGTGACACGAGCTATGAC
TCTCAACAATCTACAAACCGGCTGCCCAATGTATTCAGGCGCTTGACAGCTTCTCATC
TCTCTTCACTGAATCTCTCCAGACTGTCTGTTCTCGTTCTTCTACTCCATCAACTGATTATG
TCCAAGTTTCTCATTTATTTTAAAGCTCTCATTACGTTGGTATCATGTAAATTTGAGGAT
TGATTAAATTGAGTCTTGTGGTTTGTGAGGACTCACAATCTTCTCATTTAAAAA
AAAAA

>G577 Amino Acid Sequence (domain in AA coordinates: TBD)
MGCTASKLDSSEDAVRRCKERRRLMKDAVYARHHLAAHSDYCRSLRLTGSAALSSFAAGEP
LSVSENPVAVFLRPSSSQDAPRVPSSHSPPEPPPPPIRSKPKPTRPRRLPHILSDSSPSSS
PATSFYPTAHQNSTYSRSPSQASSVWNWENFYPPSPDSEYFERKARQNHKRPSPDYDA
ETERSDHDYCHSRDAAEEVHCSEWGDHDFRFTATSSSDGDGEVETHVSRSGIEEPPVKQ
PHQDPNGKEHSDHVTSSDCYKTKLVVRHKNLKEILDAVQDYFDKAASAGDQVSAMLEIG
RAELDRSFSKLRKTVYHSSSVFNSLSASWTSKPPLAVKYKLDASTLNDEQGLKSLCSTL
DRLLAWEKKLYEDVKAREGVKIEHEKKLSALQSQYKGGDESKLDKTKTSITRLQSLIIV
SSEAVLTSSNAILRLRDLVLPQLVELCHGLMYMWSMHEYHEIQNNIVQQVRGLINQTE
RGESTSEVHRQVTRDLESASVSLWHSSFCRIIKFQREFICSLHAWFKLSLVPLSNGDPKKQ
RPDSFALCEEWKQSLERVPDTPVASEAIKSFVNVVHVVISIKQAEVVKMKRTESAGKELEK
KASSLRISIERKYYQAYSTVGIGPGEVLDSRDPLSEKKCELAACQQRQVEDEVMRHVKA
VTRAMTLNQLTGLPNVVFQALTSFSSLFTESLQTVCSRSYSIN*

>G674 (1..786)-

ATGGTGTTTAAATCAGAAAAATCAAACCGGGAAATGAAATCAAAGGAGAAGCAAAGGAAG
GGATTATGGTCACCCGAGGAAGATGAGAAGCTTAGGAGTCATGTCTCAAATATGGCCAT
GGATGCTGGAGTACTATTCTCTTCAAGCTGGATTGCAGAGGAATGGGAAGAGTTGTAGA
TTAAGGTGGGTTAATTATTTAAGACCTGGACTTAAGAAGTCTTTATTTCACTAAACAAGAG
GAAACTATACTTCTTTCACTTCAATCCATGTTGGGTAACAAATGGTCTCAGATATCGAAA
TTCTTACCAGGAAGAACCACCAACGAGATCAAAACTATTGGCATTCTAATCTAAAGAAG
GGTGTAACTTTGAAACAACATGAAACCACAAAAAATCAACACCTTTAATCACAAAC
TCACTTGAGGCCTTGAGAGTTCAACTGAAAGATCTTCTTCATCTATCAATGTCGGAGAA
ACGCTTAATGCTCAAACCTCAAGCTTTTCGCCAAATCTCGTGTCTCGGAATGGTTAGAT

CATAGTTTGCTTATGGATCAGTCACCTCAAAAGTCTAGCTATGTTCAAAATCTTGTTTTTA
CCGGAAGAGAGAGGATTTCATTGGACCATGTGGCCCTCGTTATTTGGGAAACGACTCTTTG
CCTGATTTTCGTGCCAAATTCAGAATTTTTGTGGATGATGAGATATCATCTGAGATCGAG
TTCTGTACTTCATTTTCAGACAACTTTTTGTTCGATGGTCTCATCAACGAGCTACGACCA
ATGTAA

>G674 Amino Acid Sequence (domain in AA coordinates: 20-120)
MVFKEKSNREMKSKQKRLWSPEEDEKLRSVLYKHGHCWSTIPLQAGLQRNGKSCR
LRWVNYLRPGLKSLFTKQEETILLSLHSM LGNKWSQISKFLPGRTDNEIKNYWHSNLKK
GVTLKQHETTKKHQTP LITNSLEALQSSTERSSSSINVGETSNAQTSSFSPLV FSEWLD
HSLMDQSPQKSSYVQNLVLPBERGFIGPCGPRYLGNDSLPDFVPNSEFLLDDEISSEIE
FCTSFSDNFLFDGLINELRPM*

>G736 (1..513)
ATGGCGACTCAAGATTCTCAAGGGATTAAACTCTTTGGCAAACTATTGCATTTAACA
CGAACATAAAAAATGAAGAAGAGACACACCCCGCGAGCAAGAAGCCACAATAGCCGTT
AGATCATCATCATCATCGGATCTGACGGCCGAGAACGCTCCGATAAGATCATAGCATGT
CCAAGATGCAAGAGCATGGAGACAAAGTTCTGTTACTTCAACAATAACAAGGTAATCAG
CCTCGACACTTTTGTAAAGGCTGCCACCGTTACTGGACCGCGGTGGTGCACCTCCGGAAC
GTTCCCGTCGGCGCCGGTCGTGCGAAGTCCAAACACCTGGTCTGTCTGTGGTGGTATG
CTTGGAGATGGAAATGGTGTTCGCCAAGTCGAGCTTATAAATGGCTTGCTCGTTGAGGAG
TGGCAGCATGCCGAGCCGAGCTCACGGTAGTTTCCGGCATGATTTCCCATGAAGCGG
CTCCGGTGTACTCCGACGGTCAATCGTGCTGA

>G736 Amino Acid Sequence (domain in AA coordinates: 54-111)
MATQDSQGIKLFKTI AFNTRTIKNEEETHPPEQATI AVRSSSSDLTA EKRPDKIIAC
PRCKSMETKFCYFN NYNGNQPRHFCKGCHRYWTAGGALRNVPVGAGRRKSKPPGRVVVGM
LGDGNGVRQVELINGLLVEEWQHAAAAHGSFRHDFPMKRLRCYSDGQSC*

>G903 (96..1496)
CCCGGGTCGACCCACGCGTCCGCTCTCTCTCTGAACATACAAAAACCTACTTTTAAT
TTCTCTTCCAAGAAGTCAAGAACCAGAGAAGACATGACAAGTGAAGTTCTTCAAACAA
TCTCAAGTGGATCAGGTTTGTCTCAGCCACAGAGCTCATCAACCTGGATCATGATGAAT
CTCTCATCAATCCTCTCTTGTGAAGAAAAAGAGAAATCTCCCTGGAAATCCTGATCCGG
AAGCTGAAGTGATAGCTTTATCCCCACGACCTTGATGGCTACGAACCGGTTCTTATGTG
AGGTATGTGGCAAAAGTTTCCAAGAGAGACCAAAACTTACAGCTTTCATCGGCGAGGACATA
ATCTTCCATGGAAGTTGAAGCAGAGGACAAGCAAAGAGTGAAGAAACGTGTCTACGTTT
GCCCCGAGAAGACATGTGTCCACCATCACTCCTCTAGAGCTCTAGGCGATCTCACTGGAA
TCAAAAAGCATTTTGTCCGGAACACGGGGAGAAGAGTGGACGTGCGAGAAATGTGCTA
AGAGATACGCACTCAATCTGATTGGAAGCTCATTTCCAAGACTTGTGGTACTAGAGAGT
ACCGTTGCGATTGTGGCACCATTCTTCAAGGCGAGACAGCTTTATCACTCATAGAGCTT
TCTGCGATGCTTAGCGGAAGAAACCGCTAAGATAAACGCAGTGTCTCATCTCAACGGTT
TAGCCGCGGCTGGAGCCCGAGGATCAGTTAATCTCAACTATCAATATCTCATGGGAACAT
TCATCCCACCGCTTCAACCATTTGTACCACAACCGCAAAACAAATCCAAACCATCATCATC
AACATTTTCAGCCACCAACTTCTTCGTCGCTCTCTCTATGGATGGGACAAGATATCGCGC
CGCCTCAACCGCAACCGGACTACGATTGGGTTTTTGGAAACGCTAAGGCAGCGTCTGCTT
GCATTGATAATAATAATACTCACGATGAGCAGATTACGCAAAACGCAAAACGCAAGTTTGA
CCACTACCACTACTCTCTCTGCCCCCTTCTTATTACGAGCGACCAACCACAAAACGCAA
ACGCAAAATTCAAACGTGAATATGTCCGCGACAGCTTTACTACAGAAAGCTGCTGAAATTG
GCGCTACTTTTACAACAACCGCAGCGACCAATGACCCATCAACGTTTCTTCAAAGTTTCC
CGCTTAAATCCACCGATCAAAACCAAGTTATGACAGTGGCGAAAAGTTTTTGTCTTGT
TCGGGTCTAACAACAACATTGGGTTAATGAGTCGTAGTCATGATCATCAAGAGATCGAGA
ACGCTAGAAATGACGTTACGGTTGCGTCTGCCTTGGATGAATTACAGAATTACCCTTGGGA
AACGTAGAAGAGTTGATGGTGGAGGTGAAGTGGGTGGAGGAGGGCAAATCGGGATTTC
TCGGGGTTGGTGACAAACGTTGTGCCATCCATCGTCTATCAATGGATGGATTGAAAGA
GTTTAAAAATTTCCGGGTTAATGCATAAATTACGTAAAAGAAGAAGGAATCTTTTGTCTAT
TTCCACCATTTTCTAAGATAACATATGTATATGGTAATGAAGTTGTTTTCTTTTATTAA
TTCAATATTCTAAAACCTTATGATATATGTATAATGAATGTGTTTATCTTCAA

>G903 Amino Acid Sequence (domain in AA coordinates: 68-92)
MTSEVLQTISSGSGFAQPQSSSLDHDDES LINPPLVKKRNLPGNPDPEAEVIALSP TTL
MATNRFLCEVCGKGFQRDQNLQLHRRGHNL PWKLKQRTSKEVRKRVYVCPEKTCVHHSS

RALGDLTGIIKKHFCRKHGEKKWTCEKCAKRYAVQSDWKAHSKTCGTREYRCDCGTIFSR
DSFITHRAFCDALAEETAKINAVSHLNGLAAGAPGSVNLNYQYLMGTFIPLQPFVFPQ
QTNPNHHHQHFQPTSSSLSLWMGQDIAPPQPQPDYDWVFGNAKAASACIDNNNTHDEQI
TQANANASLT'TTTLSAPSLFSSDQPQANANANSNVNMSATALLQKAAEIGATSTTTAATND
PSTFLQSFPLKSTDQTTSDYSGEKFFALFGSNNNIGLMSRSHDHQEIENARNDVTVASAL
DELQNPWKRRRVDGGGEVGGGQTRDFLGVGVQTLCHPSSINGWI*

>G917 (32..679)

TTAGGGT'TTTAGAAAGATAGATCGATTGAAGATGAGGAAAGGTAAGAGAGTGATAAAAAA
GATAGAGGAGAAAATAAAGAGACAAGTGACATTCGCAAAGAGAAAGAAGAGTCTAATCAA
GAAGGCATATGAACCTCTCTGTCTCTGCGATGTCCACCTTGGTCTCATCATCTTCTCTCA
CTCCAAACAGGCTCTACGATTTCTGCTCCAACCTCTACCAGCATGGAGAATCTCATCATGAG
ATACCAAAGGAAAAAGAAGGTCAAACCACTGCAGAACACAGTTTCCACTCGGATCAGTG
TTCAGATTGCGTGAAGACGAAGGAATCAATGATGAGAGAGATAGAGAATCTTAAGCTGAA
TCTTCAATTGTACGACGGACATGGCTTGAATCTCTTGACCTACGACGAGCTCCTTTCTTT
TGAGCTCCATCTCGAATCTTCTCTACAACATGCTCGAGCTCGCAAGTCTGAGTTCATGCA
TCAGCAGCAGCAGCAACAAACAGATCAAAAGCTTAAGGAAAAGAAAAGGGTCAAGGAAG
CTCTTGGGAGCAGCTGATGTGGCAAGCAGAGAGACAGATGATGACGTGTCAAAGACAAAA
AGATCCTGCGCGCGCAATGAAGGAGGAGTCTCTTTTACGGTGGGGAACAAACCCACCG
ACGTTCTTACCTCTTAAGCTACCACAACCAGGCCCAAATACAGGCCCATAACTTCTCT
CTATCTATAAAAAACAAGTATAGTAAAAAGTATGACCCGTTTGGTTCGGTTATGTTG
ATACCAGACTATTAATTAACCTCGGTTAGACGTATTTACGACTTGATGCTATCTAGACCT
TTTTGCCCTTCAAAAAA

>G917 Amino Acid Sequence (conserved domain in AA coordinates:2-57)

MRKGRVKKIEEKIKRQVTFARRKSLIKKAYELSVLCDVHLGLIIFSHSNRLYDFCSN
STSMENLIMRYQKEKEGQTAEHSFHSDDQSDCVTKESMMREIENLKLNLQLYDGHGLN
LLTYDELLSFELHLESSLQHARARKSEFMHQQQQQQTDQKLKGKEKGQSSWEQLMWQAE
RQMMTCQRQKDPAPANEVGPFLRWGTTTHRRSSPP*

>G921 (116..1024)

CCAAGATCGACTCTTACTTGAATCTCTCTCAACTTTCTTCTCAGCTTACGGGAACCTC
CACACATATACATCCACAAGAACCCATATCGAAGATTCTCTACATATATTTACATGGA
TCAGTACTCATCCTCTTTGGTTCGATACTTCATTAGATCTCACTATTGGCGTTACTCGTAT
GCGAGTTGAAGAAGATCCACCGACAAGTGCTTTGGTGAAGAATTAACCGAGTTAGTGC
TGAGAACAGAAGCTCTCGGAGATGCTAACTTTGATGTGTGACAACTACAACGTCTTGAG
GAAGCAACTTATGAATATGTTAACAAGAGCAACATAACCGAGAGGGATCAAATCAGCCC
TCCCAAGAAACGCAATCCCGCGGAGAGAGGACGCATTACGCTGCGCGGTTATTGGCGG
AGTGTGCGGAGAGTAGCTCAACGGATCAAGATGAGTATTTGTGTAAGAAGCAGAGAGAAGA
GACTGTGCTGAAGGAGAAAGTCTCAAGGGTCTATTACAAGACCGAAGCTTCTGACACTAC
CCTCGTTGTGAAGATGGGTATCAATGGAGGAAATATGGACAGAAAGTGAAGTACTAGAGACAA
TCCATCTCCAAGAGCTTACTTCAAATGTGCTTGTGCTCCAAGCTGTTCTGTCAAAAAGAA
GGTTCAGAGAAGTGTGGAGGATCAGTCCGTGTTAGTTGCAACTTATGAGGGTGAACACAA
CCATCCAATGCCATCGCAGATCGATTCAAACAATGGCTTAAACCGCCACATCTCTCATGG
TGGTTACAGCTTCAACACCCGTTGCGAGCAACAGAAGAAGTAGCTTGACTGTGCCGGTGAC
TACCGTAGATATGATGAATCGAAGAAAGTGACGAGCCCAACGTCAAGAATCGATTTTCC
CCAAGTTCAGAACTTTTGGTGGAGCAAAATGGCTTCTTCTTAAACCAAGATCCTAACTT
TACAGCAGCTTTAGCAGCAGCTGTACCGGAAAATGTATCAACAGAATCATACCGAGAA
ATAGTTTAGCTTCAAATCCGTTAGAGTTTTTAGATTGAAATTTGTCATGAGTAAGAGAA
AGAGAGTAGATTATAATCCNTTGTGATACTGAAAAAAAAAAAAAAAAAAAAA

>G921 Amino Acid Sequence (domain in AA coordinates: 146-203)

MDQYSSSLVDTSLDLTIGVTRMRVEEDPPTSALVEELNRVSAENKKLSEMLTLMCDNINV
LRKQLMEYVNKSNITERDQISPPKKRKSAPAREDAFSCAVIGGVSESSSTDQDEYLCKKQR
EETVVKKEKVSRYVYKTEASDTTLVVDGYQWRKYGQKVTRDNPSPRAYFKCACAPSCSVK
KKVQRSVEDQSVLVATYEGEHNHPMPQSIDSNNGLNRHISHGGSASTPVAANRRSSLTVP
VTTVDMIESKKVTSPTSRIDFPQVQKLLVEQMASSLTKDPNFTAALAAAVTGKLYQONHT
EK*

>G922 (1..1449)

ATGGTGGCTATGTTTCAAGAAGATAATGGAACATCTTCTGTAGCTTCATCACCACCTTCAA
GTCTTCTCAACTATGTCACTCAACAGACCGACTCTCCTCGCTTCTTCATCTCCGTTTCAT

TGTCTCAAAGATCTCAAACCAGAGGAGCGTGGTCTCTACTTAATCCACCTCTTGCTAACT
TGTGCCAACCACGTGGCTTCAGGTAGCCTCCAAAACGCTAACGCAGCGCTCGAGCAGCTC
TCTCACCTCGCTTCTCCTGACGGCGACACGATGCAGCGAATCGCTGCTTACTTCACCGAA
GCGCTTGCTAACAGAATCCTTAAGTCTTGGCTGGTCTTTACAAGGCTCTTAACGCAACT
CAGACAAGAACTAACAATGTCTCTGAGGAGATTATGTTAGAAAGACTCTTCTTTGAGATG
TTCCCCGATACTCAAAGTCTCTTACTTGCTCACTAATCGAGCTATACTCGAGGCTATGGAA
GGAGAGAAGATGGTTCATGTGATTGATCTCGATGCTTCTGAGCCAGCTCAATGGCTTGCT
TTGCTTCAAGCTTTTAACTCTAGGCCTGAAGGTCCACCTCATTTGAGAATCACTGGTGT
CATCACCAGAAGGAAGTGCTTGAACAAATGGCTCATAGACTCATTGAGGAAGCAGAGAAA
CTCGATATCCCGTTTTCAGTTTAAATCCCGTTGTGAGTAGGTTAGACTGTTTAAATGTAGAA
CAGTTGGGGTTTAAACAGGAGAGGCCTTAGCCGTTAGCTCGGTTCTTCAATTGCATACC
TTCTTGGCCTCTGATGATGATCTCATGAGAAAGAACTGCGCTTTACGGTTTTCAGAACAAAC
CCTAGTGAGATTGACTTGAGAGAGTTCTAATGATGAGCCATGGCTCTGCAGCTGAGGCA
CGTGAGAATGATATGAGTAACAACAATGGGTATAGCCCTAGCGGTGACTCGGCTCATCT
TTGCTTTTACCAAGTTTCAAGGAGGACTGATAGCTTCTCAATGCTATTTGGGGTTTGTCT
CCAAAGGTTCATGGTGGTCACTGAGCAAGACTCAGACCACAACGGCTCCACACTAATGGAG
AGGCTATTAGAATCACTTTACACCTACGCAGCATTGTTTGATTGCTTGAAACAAAAGTT
CCAAGAACGCTCAAGATAGGATCAAAGTGGAGAAGATGCTCTTCGGGGAGGAGATCAAG
AACATCATATCTCGAGGGATTGAGAGAAGAGAAAGACACGAGAAGCTTGAGAAATGG
AGCCAGAGGATCGATTGGCTGGTTTGGGAATGTTCTCTTAGCTATTATGCGATGTTG
CAGGCTAGGAGATTGCTTCAAGGGTGGGTTTGTATGGGTATAGAATCAAGGAAGAGAGC
GGGTGCGCAGTAATTTGCTGGCAAGATCGACCTCTATACTCGGTATCAGCTTGAGATGC
AGGAAGTGA

>G922 Amino Acid Sequence (conserved domain in AA coordinates:225-242)

MVAMFQEDNGTSSVASSPLQVFSTMSLNRPDLLASSSPFHCLKDLKPEERGLYLIHLLLT
CANHVASGSLQANANAALQLSHLASPDGDTMQRIAAYFTEALANRILKSWPGLYKALNAT
QTRTNVSEIEIHVRLFFEMFPILKVSYLLTNRAILEAMEGEKMHVVIDLDASEPAQWLA
LLQAFNSRPEGPPLRLITGVHHQKEVLEQMAHRLIEAEKLDIPFQFNPVVSRLDCLNVE
QLRVKTGEALAVSSVLQLHTFLASDDLMRKNCALRFQNNPSGVLDLQRLVLMMSHGSAAEA
RENDMSNNNGYSPSGDSASSLPLPSSGRTDSFLNAIWGLSPKVMVTEQSDSHNGSTLME
RLLESLYTYAALFDCLTKVPRTSQDRIKVEKMLFGEEIKNIISCEGFERRERHEKLEKW
SRIDLAFGNVPLSYAMQLQARLLQCGFDGYRIKEESGCAVICWQDRPLYSVSAWRC
RK*

>G932 (206..1213)

CCACGCGTCCGACCACTTGTACCTCTTTGTCTTAAGTACTCTTTAACCTACAATTCCT
AAGCTCTCAAGCCACAAAAAACCACAAACCGTTCTTCAACCAATATATATATCTGATCATC
ATCAAAGTCCTTCTCTCTGCTCATACCACAAACCGTTCCATTCTTCCCCTAATCACAAAG
TGATATTACATAGAGAAGATAGAGATGGGAAGACCACCATGCTGTGACAAGATTGGAGT
GAAGAAAGGACCATGGACACCAGAGGAAGATATCATCTTGGTTTCTTACATCCAAGAAC
TGGTCCTGGAAACTGGAGATCTGTGCCTACTCACACAGGTTTGAGGAGATGTAGCAAAAG
CTGTAGATTGAGGTGGACTAATTATCTTCGACCTGGGATCAAGCGTGGAAATTTACCCGA
GCATGAAGAGAAGATGATCTCCATCTTCAAGCTCTTTGGGAAACAGGTGGGCAGCTAT
AGCATCATATCTTCCAGAAAGGACAGACAATGATATAAAGAACTATTGGAACACTCATTT
GAAGAAAAAGCTCAAGAAGATGAATGATTCTTGTGATAGTACTATCAACAATGGCCTTGA
TAATAAAGACTTCTCCATATCAAACAAAAACACTACCTCACATCAAAGCAGCAACTCCAG
TAAAGGTCAATGGGAGAGAAGGCTTCAGACAGATATCAACATGGCTAAACAAGCTCTTGT
TGATGCCCTTGTCTATTGACAAACCAACCAAAACCACTAATTTTTCTATTCCCGATCTTGG
TTATGGTCCATCATCTTCTCGTCTCTACCACCACCACCACCACCACCACCACGAG
AAACACTAATCCATACCCATCTGGGGTCTATGCTTCAAGTGCTGAGAACATTGCTCGTTT
GCTTCAGAAATTTTATGAAGACACACCAAGACCTCGGTGCCCTTGCCGTTGACGCCAC
CGAGATGGCTATCACCACGGCAGCTTCGAGCCCTAGCAACCGAAGGAGACGGAGAAGG
GATTGACCAATCTTTGTTTCAGCTTCAACTCCATAGATGAAGCTGAAGAGAAGCCTAACT
AATAGACCATGACATTATGGTCTAATTACACAAGGCTCTCTTCTTTGTTTCGAGAAATG
GCTCTTTGATGAGCAAGGCCACGATATGATCATCAATAACATGTCATAGAGGGTCAGGA
AGTGTGTTGTTAGAAAGCATTAAAGTTTGACGATTGCTTGAGGAACCACGAGGCTTAGT
TATAACAATTTGTATAATTAAGTACTCTTAGTTTTGTTTTCAATCCTTATTATGATCA
TATTGCAGTAATTAGGGATTTTAGTCTTAGTAGTAACTCTTAAGTTTTAACACATTTTT

CTCTATCTTTTATAGTAGTAACCTCTTTATTTTCTTAAATCTTTGTCGACGTGGAGATG
ATATCTTCTATGTAGTAGAACTCAAAGTGTACATCATCTTTATTAATGTAACGTCTTT
TAAAAAAAAAAAAAAAAA

>G932 Amino Acid Sequence (domain in AA coordinates: 12-118)
MGRPPCCDKIGVKKGPWTPPEEDIILVSYIQEHGPGNWRVPTHTGLRRCCKSKSCLRWITNY
LRPGIKRGNFTEHEEKMILHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKLKKMN
DSCDSTINGLDNKDFSISNKNTTSHQSSNSSKGQWERRLQTDINMAKQALCDALSIDKP
QNPTNFSIPDLGYGPSSSSSSSTTTTTTTTTTRNTNPYPGVSASSAENIARLLQNFMKDT
PKTSVPLPVAATEMAITTAASSPSTTEGDGEGIDHSLFSFNIDEAEKPKLIDHDINGL
ITQGSLSLFEKWLFDQSHDMIINNMSLEGQEVLF*

>G599 (152..1579)

TCGACAGAACAGCTTCGTTGTCACTTGTCACTTCTATAAATCGCATCCCCATTGACAACTT
TTCATTCCATCAAACTCTCTCTCTATATCTCTCTCTCTATATCTCTCTCTATATCT
CTCTCTCTCTCACTCTCTCTTTCTTTCAAATGGAAAACTCATGGTTCCGACATGGAG
ACCCGACCCGGTTTACCGTCCACCGGAAACACCACTCGAACCGATGGAGTTTTAGCTCG
TTTCATGGAGCGTCTCTGCTCTCGAAGTCTCAAGGCTCTAACACCAACCAACCTCAGAT
TCTCTCTCTCAAAACCGAAGAAGAAGAAGAAGAAGAACCCATCTCTCTGTCTAGACGG
CGACGGCGACACGGAAGACACCGGACTTGTCAACGGAAACCCATTCTCTCTGCTTGTTC
AGAACTTCTCAAATGGTTCATGGATCGTATCTTGTCTCACTCTCAAGAAGTATCACCAAG
AACATCTGGTCCGCTATCTCAGAGTAGTGGTCCACTTAATGGTTCTTTGACCGACAGTCC
TCCTGTGTCTCTCCGAATCCGACGACATTAAGCAATTTTGCAGAGCGAACAATAATTC
ATTGAACAGTGTAAATTTCTCAGTTCCGTTCAACGGCGGCAACTCCGGGACCTATAACCGC
TAGCATACACAGTCCAAAGACGGTGGGACGGTGGCTTAAGGACCGGAGAGAGAAAAAGAA
AGAGGAGACTCGGGCTCATAACGCTCAGATTACGCTGCTGTCTGTCTGCGCGGCTTGC
TGCAGCTGTGCTGCTATTGTCAGCAGCCACCGTCCGCTCTTAGCTGTGGTAAGGATGA
GCAGATGGCTAAAACTGACATGGCCGTTGCTTCTGCTGCGACCCCTTGTGGCTGCTCAGTG
TGTGGAAGCTGCTGAAGTTATGGGAGCTGAGAGAGATATTTGGCTTCTGTTGTTAGCTC
CGCCGTCAATGTTCTGCTGCGGAGATATTATGACTCTCACCGCCGGAGCAGCTACAGC
TTTAAGAGGAGTGCAACATTTGAAGGCAAGGGCAATGAAGGAAGTGTGGAACATAGCATC
AGTGATACCAATGGATAAAGGACTCACCTTCTACAGGAGGAAGCAGCAATAATGTTAATGG
TAGCAATGGAAGCTCAAGCAGTAGTCACAGTGGTGAACCTGTACAAACAGGAGAATTTCTT
GGGAACCTGTAGTAGAGAAATGGCTCGCTAGAGGTTGTGAACCTCTCAAACGCACTCGCAA
AGGTGATCTCCACTGGAAGATAGTATCTGTTACATCAACAAAATGAATCAGGTTATGTT
GAAGATGAAGAGCAGGCATGTTGGAGGAACCTTCACCAAGAAGAAAAAGAACATTGTGCT
TGATGTGATCAAGAATGTCCCGGCTGGCCTGGACGACATTTGCTAGAGGGAGGAGATGA
CTAAGATACCTTCGGTTGAAGACGGTTATGCGAGGTGATGTTGAATTCGAGGTCAAGAG
CCAAAGGAATATGAAATGTGGACACAAGGTGTCTCAAGGCTTCTGTTCTTGCTGCTGA
GAGGAAGTTTAGGATGTGAATAAACGTTCAATGGCTGCTTGGTTTAAGTGTGAGTTTTTT
TTTAACCTATGTGGTCAAATTTTATTAGTAGGGGTTCTTTTAAGGTAATGGTTTTTTGGG
TTGGGTATAGGATAAAATGGACCTACCAAGTCAAGGTGAGGAAGCATTTGGGTAAACAAAA
CTTAGTGGGGGTGATCTGTAATATCTATGTTCTTAGTTTTTTTTTGGTTGTTGGTGGTCT
TTTTGTATAAAAAACAAAGTTGAAGTAATAGATATATAGTATGTTTTAATTTTAAA

>G599 Amino Acid Sequence (domain in AA coordinates: 187-219, 264-300)

MEKLMVPTWRPDPVYRPPETPLEPMEFLARSWSVSALEVSKALTPPNPQILLSKTEEEEE
EPISSVVDGDDTETDGLVTGNPFSFACSETSQMVMRILSHSQEVSPRTSGRLSHSSG
PLNGSLTDSPPVSPPEDDIKQFCRANKNSLNSVNSQFRSTAATPGPITATATQSKTVGR
WLKDRREKKKEETRAHNAQIHAASVAVGAAVAAIAAATAASSSCGKDEQMAKTDMAVA
SAATLVAAQCVEAAEVMGAEREYLASVVSSAVNVRSAGDIMTLTLAGAATALRGVQTLKAR
AMKEVWNIA SVIPMDKGLTSTGGSSNNVNGSNGSSSSSHSGELVQENFLGTCSREWLR
GCELLKTRKGDHLHWKIVSVYINKMNQVLMKMSRHVGGTFTKKKKNI VLDVIKNVPAPW
GRHLLGGDDLRYFGLKTVMRGDVEFEVKSQREYEMWTQGVSRLLVLAERKFRN*

>G804 (114..1139)

ATACTCCAAGAATTTATAGGTTATAAGTAAAAATTCAGTACAAGTTGTTTGTGTTGTTA
TTCCATTTTCTTGTGTGTTTTTTTCCCATTAATTTATAAATTTTATAAGCAATATGGAGT
CCCACAACAACACAGAGCAACAACACCACTGGTTCCGGCCCATCTGGTCCCATCCA
TGGGACCAATCTCCGGTTCAGTCTCATTAAACCACTGCTCCAACTCCACTACCACCA
CCGTCAACCGCCGCTAAACACCCGCAAAACGACCGTCCAAGGACCGTCACATCAAAGTAG

ACGGACGTGGCCGGAGGATACGTATGCCGGCTATCTGCGCAGCACGTGTCTTCCAATAA
CACGTGAGTTACAACACAAATCGGACGGCGAGACTATAGAGTGGCTGCTCCAACAAGCGG
AGCCAGCTATCATCGCAGCCACCGGAACCTGGAACCATACCGGCGAATATCTCTACTTTGA
ACATCTCTCTTGAAGCAGTGGCTCTACTCTTTCAGCTCCACTGTCTAAATCTTTCCACA
TGGGAAGAGCGGCTCAAAACGCTGCCGTTTTTGGGTTCAGCAACAGCTTTATCATCCTC
ATCATATCACGACAGATTCTTCTTCTTCTTCTTCTTCCAAAACATTCCGTGAAGAAGATC
TTTTTAAAGATCCTAATTTTCTAGATCAAGAACCGGTTCAAGATCACCTAAACCGGGAT
CCGAAGCTCCTGATCAAGATCCGGGTTGACCCGGTCAAGAACACAAAATATGATACCGC
CGATGTGGGCACTAGCGCCAACGCCAGCCTCCACAAACGGAGGTAGTGCTTTTTGGATGT
TACCAGTCGGAGGAGGAGGAGGTCCGGCTAACGTTCCAGGATCCATCACAGCACATGTGGG
CGTTTAAATCCGGGTCATTACCCGGGTCGAATCGGGTCGGTTACGTAGGGTCTATGTAG
TGGGAGGTCAACAGTTAGGGTTAGGTGTTCAGAAAATAACAATTTGGGGCTATTTCCG
GCGGAGGAGGAGACGGTGGTCCGGTTCGGAATGAGTCTTGAGCAAAAGCCTCAAC
ATCAAGTGAGTGATCATGCTACTAGAGACCAAAATCCTACTATAGATGGTTCTCCTGAA
AGACTTCATGATTTCTTTGGTTTTTAAAAGTGTGAATGTGTGATTTATTGCAACTTTG
TTGAGGACTCCAATGTTAATATGGGTTTTAGGGTTGGCTTTTCGGGATTGCCAAATTGTT
ATT

>G804 Amino Acid Sequence (domain in AA coordinates: 54-117)
MESHNNNQSNNTTGS AHLVPSMGPISGSVSLTTTAPNSTTTT VTAAKTPAKRPSKDRHI
KVDGRGRRI RIMPAICARVFQLTRELOHKSDGETIEWLLQQAEP AIIAATGTGTIPANIS
TLNISLRSSGSL SAPLSKSFHMGRAAQNAAVFGFQQQLYHPHHITDSSSSSLPKTFRE
EDLFKDPNFLDQEPGSRSPKPGSEAPDQDPGSTRSRTQNMIPPMWALAPTASTNGGSAF
WMLPVGGGGGPANVQDPSQHMWAFNPGHYPRIGSVQLGSM LVGGQQLGLGVAENNNLGL
FSGGGDGGRVGLGMSLEQKPQHVS DHATRDQNPTIDGSP*

>G1062 (297..1781)
CAAAAAAAAAAGTTTCAATTTTGAAGCTCTGAGAAATGAAATCTATCATTCTCTCTCTC
TATCTCTATCTTCTTTTTCAGATTTTCGTTCTTCAATTCATGAAATCCTCGTGATTCTAC
TTTAATGCTTCTCTTTTTTACTTTTCCAAGTCTCTGAATATCAAAGTATATATCTTTT
GTTTTCAAACCTTTTGCAGAATGTCTTCAAGCTTCCAAATTCAGTTAAAGGTCTCAACT
TTGCAGAAATTTCTCTAAAGGTTTCAGACTTTGGGGTAAAGGTGTCAACTTTGGCGATGG
GTCTTGACGGAAACAATGGTGGAGGGGTTGGTTAAACGGTGGTGGTGGAGAAAGGAAG
AGAACCAGGAAGGTTTCATGGGAAGGAATCAAGAAGATGGTTCTTCTCAGTTTAAGCCTA
TGCTTGAAGGTGATTGGTTTTAGTAGTAACCAACCATCCACAAGATCTTCAGATGTTAC
AGAATCAGCCAGATTTTCAGATACTTTGGTGGTTTTCTTTTAAACCTAATGATAATCTTC
TTCTTCAACACTCTATTGATCTTCTTCTTCTTGTCTCTTCTCAAGCTTTTAGTCTTG
ACCTTCTCAGCAAAATCAGTTCTGTCAACTAACAACAACAGGGTTGTCTTCTCAATG
TTCTTCTTCTGCAACCCCTTTTGATAATGCTTTTGAGTTTGGCTCTGAATCTGGTTTTTC
TTAACCAAATCCATGCTCCTATTTTCGATGGGGTTTGGTTCTTTGACACAATTGGGGAACA
GGGATTTGAGTTCTGTCTGATTCTTGTCTGTCTGCTCGGTCACTTCTTGCGCCGGAAGCA
ACAACAACAACAACAATGTTGTGTGGTTTTTCACAGCTCCGTTGGAGTTGGAAGGTTTTG
GTAGTCTGTCTAATGGTGGTTTTGTGGGAACAGAGCGAAAGTTCTGAAGCCTTAGAGG
TGTTAGCATCGTCTGGTGCACAGCCTACTCTGTTCCAGAAACGTGCAGCTATGCGTCAGA
GCTCTGGAAGCAAAATGGGAAATTCGGAGAGTTCCGGGAATGAGGAGGTTTAGTGATGATG
GAGATATGGATGAGACTGGGATTGAGGTTTCTGGGTTGAACTATGAGTCTGATGAGATAA
ATGAGAGCGGTAAAGCGGCTGAGAGTGTTCAGATTGGAGGAGGAGGAAAGGTAAGAAGA
AAGGTATGCCTGCTAAGAATCTGATGGCTGAGAGGAGAAGGAGGAAGAAGCTTAATGATA
GGCTTTATATGCTTAGATCAGTTGTCCCCAAGATCAGCAAAATGGATAGAGCATCAATAC
TTGGAGATGCAATTGATTATCTGAAGGAACCTTACAAAGGATCAATGATCTTCACAATG
AACTTGAGTCAACTCCTCTGGATCTTTGCCTCCAATTCATCAAGCTTCCATCCGTGGA
CACCTACACCGCAAACCTTTTCTTGTGCTGTCAGGAAGAGTTGTGTCCCTCTTCTTTAC
CAAGTCCATAAGGCCAGCAAGCTAGAGTTGAGGTTAGATTAAGGAAGGAAGAGCAGTGA
ACATTATATGTTCTGTGGTTCGTAGACCGGGTCTGTTGCTCGCTACCATGAAAGCTTTGG
ATAATCTTGGATGGATGTTTCAGCAAGCTGTGATCAGCTGTTTTAATGGGTTTGCCTTGG
ATGTTTTCCGCGCTGAGCAATGCCAAGAAGGACAAGAGATACTGCCTGATCAAATCAAAG
CAGTGCTTTTTCGATACAGCAGGGTATGCTGGTATGATCTGATCTGATCCTGACTTCGAGT
CCATTAAGCATCTGTTGAAGCAGAGCTAGAAGAATAAGTCCCTTTAAATCTGCAATTTT
CTTCTCAACTTTTTTTCTTATGTATCAACTTCAATCTAAGCATGTAATGCAATTGCAAAAT

GAGAGTTGTTTTTAAATTAAGCTTTTGAGAACTTGAGGTGTGTGTTGTTGGATACATAAC
TTCAACCTTTTATTAGCAATGTTAACTTCCATTATGTTTCATCTT
>G1062 Amino Acid Sequence (domain in AA coordinates: 308-359)
MGLDGNNGGGVWLNGGGGERENEESGWRNQEDGSSQFKPMLEGDWFSNQPHQPDLQM
LQNQPDFRYPFGGFFNPNDNLLQHSIDSSSSCSPSQAFSLDPSQQNQFLSTNNNKGCLL
NVSSANFPDNAFEFGSESGFLNQIHAPISMGFGSLTQLGNRDLSSVPDFLSARSLLAPE
SNNNTMLCGGFTAPLELESGFGSPANGFVGNRAKVLPLEVLAASSGAQPTLFQKRAAMR
QSSGSKMGNSESSGMRRFSDDGDMDETGIEVSGLNYESEDINESGKAAESVQIGGGGKGK
KKGMPAKNLMAERRRRKKLNDRLYMLRSVVPKISKMDRASILGDAIDYLKELLQRINDLH
NELESTPPGSLPPTSSSFHPLTPTPQTLSCRVKEELCPSSLSPKPGQQARVEVRLREGRA
VNIHMFCCRPGLLLATMKALDNLGLDVLVQQAIVISCFNGFALDVFRAEQCQEGQEILPDQI
KAVLFDTAGYAGMI*
>G1322 (213..833)

>G1322 (213..833)

AAAGTTATTGATAGTTTCTGTTACTTTATTAATTTTTTAAGGTTATGTGTATTATTACCAAT
TGGAGGACTATATAGTCGCAAGTCTCAACCTATAAAAGAAAACATTCGTCGATCATCTT
CCCGCCTCGAGTATCTCTCTCTCTCTCTCTCTCTGTTTTCTTTATTGATTGCATAGA
CAAAAATACACACATACACAACAGAAAGAAAGATGGAGACGACGATGAAGAAGAAAGGGA
GAGTGAAGGCGACAATAACGTCACAGAAAGAAAGAAGGAACAGTGAGAAAAGGACCTT
GGACTATGAAGAAGATTTCATCTCTCTTTAATTACATCTCTTAATCTGGTGAAGGTCTTT
GGAACTCTGTGCGCAAAGCCTCTGGTCTTAAACGTCATCGAAAAAGTTGTGCGGCTCCGGT
GGCTGAACTATCTCCGACCAGATGTGCGGCGAGGGAACATAACCGAAGAAGAACAGCTTT
TGATCATTCAGCTTCATGCTAAGCTTGGAAACAGGTGGTGAAGATTGCGAAGCATCTTC
CGGGAAGAACGGACAACGAGATAAAGAACTTCTGGAGGACAAAGATTGAGAGACACATGA
AAGTGTCTATCGGAAAATATGATGAATCATCAACATCATTGTTCCGGGAAACTCACAGAGCT
CGGGGATGACGACGCAAGGCAGCTCCGGCAAAGCCATAGACACGGCTGAGAGCTTCTCTC
AGGCGAAGACGACGACGTTTAATGTGGTGGAACAACAGTCAAACGAGAATTACTGGAACG
TTGAAGATCTGTGGCCCGTCCAATTGCTTTAATGGTGACCACCATGTGATTTAAGATATAT
ATATAGACCTCCTATACATTATATGATCCCCAGCTGGGTTTTTTTGTATGGTACGTTATTT
GGTTTTTCTATTGCTGAAATGTGCTGTCATTTAATTTACATACGAAAGATGCATTAATC
ATTAATCTTCAATACATATGGAGGTGGTGTGTTGAGTAAAAAAAAAAAAAAAA

>G1322 Amino Acid Sequence (domain in AA coordinates:26-130)
METTTMKKGRVKATITSQKEEGTVRKGPWTMEEDFILFNYYILNHGGLWNSVAKASGLK
RTGKSCRLRLWNLRLPDVRRGNHNTIEEQLLIIQLHAKLGNRWSKIATKHLPGRTDNEIKNF
WRTKIQRHMKVSSSENMMNHQHCHSGNSQSSGMMTTQGSNGKAIDTAESFSQAKTTTFNVVE
QSQSNGYVNVDELWPVHLLNGDHHVI*

>G1331 (1..786)
ATGGTGGAGAAGTTTGGAGAAAGGGTCCATGGACCGCCGAAGAAGACCGTCTTTTGATC
GAATACGTCCGTGTTACCGGTGAAGGTCGTTGGAACTCTGTCTCTAAACTCGCAGGATTG
AAAAGGAATGGCAAAAGCTGCAGACTAAGATGGGTGAATTACCTTAGACAGACCTCAAG
AGAGGACAGACTACCTCCACATGAAGAAAGTAACTTGAGCTACACGCTAAGTGGGGGA
AATAGGTGGTCAACAATTGCACGTAGTTTACCAGGAAGAAGACAGACAATGAGATCAAGAAG
TATTGGAGAAGCCATTTCAGAAAAAGGCCAAGCCTACGACTAACAAATGCGGAGAAGATA
AAGAGTCGTCTCCTAAAAAGGCAACACTTCAAGGAACAGAGAGAAATAGAGTTGCAACAA
GAACAGCAGTGTGTTTCAGTTTCGACCAACTCGGTATGAAAAAGATCATCTCTTTACTCGAA
GAAAACAATAGCAGTAGTACGATGCGGTGGTGATGTGTTCTATTATCTGATCAA
ATAACACATTTCATCAAAACCCCTTTGGCTATAACTCTAATTCATTAGAGGAGCAGTTACAA
GGTAGATTTTCTCCTGTAAACATACCTGATGCTAATACTATGAACGAAGACAATGCCATA
TGGGACCGGGTTTTGGAACATGGATGTTGTAAATGGACATGGTGGGAACCTGGGTGTTGTG
GCTGCTACTGCTGCTTGTGGCCCCAAGGAAGCCCTATTTCCTAACTTGGTGATTCATT
TGTTAA

>G1331 Amino Acid Sequence (conserved domain in AA coordinates:8-109)
MVEEVWRKGPWTAEDRLIIYVRVHGEGRWNSVSKLAGLKRNGKSCRLRWVNYLRPDLK
RGQITPHEESIILELHAKWGNRWSTIARSLPGRDTDNEIKNYWRTHFKKKAKPTTNNAEKI
KSRLLLKRQHFKEQREIELQEQQLFQFDQLGMKKIISLLEENNSSSSSDGGGDVFYYPDQ
ITHSSKPFYGNSNSLEEQLQGRFSPVNIPDANTMNEDNAIWDGFWNMDVVGNGHGGNLGVV
AATAACGPRKPYFHNLVIPFC*
>G1521 (1..891)

ATGCCTCCATTACCGTCCTCCACGGCGCCTTCGTCTTCGAGACATCTTCGATCGCCGGAA
AGTATCGCGAAATTTGCAGGGAGAGCAATATTTCTGCTTTACAGGGGAAATCGTGTCGG
ATATGCCTCGAAAATCTAACCGAGCGAAGATCCGCCGCCGTGATCACGGTGTGCAAGCAC
GGATACTGCCTTGCTTTGATTTCGGAAGTGGAGCAGCTTCAAGAGGAATTGTCCTCTTTGT
AACACTCGTTTTGATTCTCGTTTATCGTTAGTGATTTGCTTCTAGAAAATACCATAAG
GAGCAATTACCAATTCTTCGTGATCGTGAGACTTTAACTTATCATCGGAATAATCCTTCC
GATCGCCGGAGGATAATTCAAAGGTCGAGGGATGTTTTGGAAAACCTAGCTCAAGATCA
AGGCCATTGCCATGGCGGAGATCATTGACGACAGGTTTCACTTCTGATTCTGTTATC
TTCCAGCGAAAGCTTCAGTGGCGAGCTAGCATATACACTAAGCAATTACGAGCTGTTCCGA
TTACATTCAAGGCGCTTGGAACTAAGTTTGCGCGTGAATGATTACACCAAAGCAAAGATA
ACTGAAAGAATTGAGCCATGGATTAGAAGAGAGCTTCAGGCAGTCCCTGGAGATCCTGAT
CCCTCAGTTATTGTTTCATTTCGCTCAGCTCTTTTCATCAAAGGCTTGAGAGAGAGAAT
AATCGACAAAACCGGGCAGACCGGGATGTTGGTGGAAGATGAAGTCTCTCTCTTCGAAAA
TTCTTGCTGATAAGGTGGATATATTTTGGCATGAACCTAAGATGTTTTCGGGAGAGTATA
CTCAGGATGGAGACTTATGATGCAGTGGTTGAATACAATGAGGTGGAGTAA

>G1521 Amino Acid Sequence (domain in AA coordinates: 39-80)
MPPLPSSSTAPSSSRHLRSPESIAKFAGRAIFPALQKSCPICLENLTERRSAAVITVCKH
GYCLACIRKWSSFKNRNCPLNTRFDSWFIVSDFASRKYHKEQLPILRDRELTLYHRNNPS
DRRRIIQSRDVLNSSSRSLPWRRSFGRPGSVDPDSVIFQRKLQWRASIYTKQLRAVR
LHSRRLELSLAVNDYTKAKITERIEPWIRRELQAVLGDPPSVIVHFASALFIKRLEREN
NRQTGQTGMLVEDEVSSLRKFLSDKVDIFWHELRCFAESILTMETYDAVVEYNEVE*

>G183 (1..1458)

ATGAGTGATTTTGATGAAAACCTTCATCGAAATGACGTCGTATTGGGCTCCACCATCCAGT
CCTAGCCCAAGAACGATATTGGCAATGCTGGAGCAAACCGACAATGGTCTGAATCCAATC
AGTGAGATCTTCCCTCAAGAAAGCTTGCCAAGAGATCATACTGATCAATCTGGACAAAGA
TCTGGTCTTCGTGAGAGACTGGCTGCAAGAGTAGGATTCAATCTTCCAACACTCAATACA
GAAGAAAACATGAGTCTCTTTGGATGCATTTTTTCAGGAGCTCGAATGTTCTTAATTCTCCT
GTCGTTGCAATCTCTCCAGGATTCAGTCCATCAGCACTATTGCATACTCCCAATATGGTC
AGTGATTCTTCCCAGATTATCCCTCCGTCTTCAGCCACCAATTACGGACCTCTAGAGATG
GTGGAAACTTCCGGTGAAGACAATGCAGCGATGATGATGTTCAACAACGATCTTCTTTAT
CAGCCGTACAATGTTGATCTGCCCTTCTCTAGAAGTCTTTGATGATATTGCAACGGAAGAG
TCCTTTTATATCCCATCTTATGAACCTCATGTTGACCAATTGGAACCTCTTTAGTCACA
TCCTTTGAATCTGAACCTGTTGACGATGCCCATACCGACATCATCTCCATTGAGGACAGT
GAGAGCGAGGATGGAAACAAAGATGATGACGACGAGGACTTCCAATACGAAGACGAAGAC
GAAGACCAATACGACCAAGATCAAGATGTAGATGAAGATGAAGAGGAAGAAAAAGATGAA
GACAATGTTGCATTAGATGATCCTCAACCTCCACCTCCAAAGAGAAGGAGATATGAGGTA
TCAAACATGATTGAGCCACAAGAACAAGCAAGACACAAAGGATCATACTTCAGATGGAA
AGCGACGAAGACAATCCTAACGATGGTTATCGCTGGAGAAAATACGGTCAGAAAAGTCGTC
AAAGGAAATCCTAATCCGAGGAGTACTTCAAGTGACACAAACATCGAGTGCAGAGTGAAA
AAACATGTGGAGAGGAGCAGACAATATCAAGTTGGTTGTGACTACATACGATGGGATA
CACAAACCATCCTTCACCACCTGCACGTAGAAGCAATTCCAGTTCAAGGAACCGGTCTGCA
GGGGCAACAATACCTCAAATCAGAATGATCGAACCAGTCGGTTAGGTAGGGCTCCTCCT
ACTCCTACTCCTCCTACTCCTCCTCCTTCTGCTTACACACCTGAGGAGATGAGGCCTTTC
TCTTCGTTGGCTACAGAAATTGATCTGACAGAGGTTTATATGACCGGAATCTCTATGCTG
CCGAATATACCGGTTTACGAGAATTCGGGTTTTATGTACCAGAATGATGAACCGACGATG
AATGCGATGCCGGATGGTTTCAGATGTGTACGATGGGATCATGGAACGCCTGTATTTAAG
TTTGGTGTGACATGTAG

>G183 Amino Acid Sequence (domain in AA coordinates: TBD)
MSDFDENFIEMTSYWAPPSSPSRPTILAMLEQTDNGLNPISIEIFPQESLPRDHTDQSGQR
SGLRERLAARVGFNLPTLNTEENMSPLDAFFRSSNVPNSPVVAISPGFSPSALLHTPNMV
SDSSQIIPSSATNYGPLEMVETSGEDNAAMMMFNNDLPYQPYNVLDPSLEVFDIATEE
SFYIPSYEHPVDPIGTPLVTSFESELVDDAHTDIISIEDSESEDGKNKDDDDDEFQYEDD
EDQYDQDQDVEDEEEEDNVALDDPQPPPPKRRRYEVSNMIGATRTSKTQRIILQME
SDEDNPNDGYRWRKYGQKVVKGNPNPRSYFKCTNIECRVKKHVERGADNIKLVVTTYDGI
HNHPSPPARRSNSSSRNRSAGATIPQNQNDRTSRLGRAPPTPTPTPPSSSYTPEEMRPF
SSLATEIDLTEVYMTGISMLPNIPVYENSGFMYQNDEPTMNPMPDGSVDYDGIMERLYFK
FGVDM*

>G2555 (177..956)

CTGTTTTTGTATCCGTGTAAATTAATCACACGGTAGTTTTTGTATGAAAAGACAACAATCG
GAGAACAATCTGGTCTGCTGCTAAAAATTAATAAATTGTTTTGTCTAATTGTCTCCACCC
ATAAAAAAGCGCAATTCAATTCACCGACTAAAGACATTCTCCGGTGGAGACCCCGATGC
AATCCACTCATATAAGCGGCGGAAGTAGCGGTGGTGGTGGTGGAGGAGGAGAGGTGA
GTCGAAGTGGATTATCTCGATCCGTTACGCTCCAGCTACTTGGATTGAAACCCTACTCG
AAGAAGATGAAGAAGAAGGTTTAAACCTAACCTTTGTTAACAGAGCTGCTTACTGGTA
ATAATAACTCTGGAGGAGTGATAACGAGTCGTGACGACTCGTTCGAGTTCCTGAGTTCG
TTGAGCAAGGATTGTATAATCATCATCAAGGTGGTGGCTTTCACCGTCAGAATAGTTCTC
CGGCTGATTTTCTTAGTGGGTCTGGTTCTGGGACTGATGGGTATTTCTCTAATTTTGGTA
TTCCGGCGAATTATGACTATTTGTGCGACCAACGTTGATATTTCTCCGACTAAACGGTCTA
GAGATATGGAAACACAGTTTTCTTCTCAGCTGAAAGAAGAGCAAATGAGTGGTGGGATAT
CAGGAATGATGGATATGAACATGGACAAGATTTTTGAGGATTGAGTTCCTTGTAGGGTTC
GTGCTAAACGTGGTTGTGCTACTCATCCTCGTAGCATTGCTGAACGGGTGAGAAGAAGCG
GAATAAGTGATCGGATTAGGAGGCTGCAAGAGCTTGTTCCTAACATGGATAAGCAAACCA
ACACTGCAGACATGTTGGAAGAAGCTGTGGAGTATGTGAAGGCTCTTCAAAGCCAGATCC
AGGAATTGACAGAGCAGCAGAAGAGATGCAAATGCAAACCTAAAGAAGAACAATAATGTA
TCCTTTAGGATTGATATATCTGTATTTTATTTTTGTACTATCTAAAAATGGTGATGATC
TGTTTCGAAAATTCGAAACATGATCTTATATATTGAACTAGAAAAAATAGATATATATGAA
TTTTAGCTGTAAAATTTTTGTACAATAAGGAGAAAAAGATTTAGAAGAGTCAATAAAAAAG
ATGATGTTTACAAGTCAAAAAA

>G2555 Amino Acid Sequence (domain in AA coordinates: 175-245)

MQSTHISGGSSGGGGGGEVSRSGLSRIRSA PATWIETLLEDEEEGLKPNLCLTELLT
GNNNSGGVITSRDDSTFEFLSSVEQGLYNHHQGGGFHRQNSSPADFLSGSGSGTDGYFSNF
GIPANYDYLSNVDISPTKRSRDMETQFSSQLKEEQMSGGISGMMDMNDKIFEDSVPCR
VRAKRG CATHPRISIAERVRRTRISDRIRRLQELVPNMDKQNTADMLEEAVEYVKALQSQ
IQELTEQQKRCCKPKEEQ*

>G375 (53..1171)

TCGACAAAACTCTCACTCTCCCTCAAAC TAAACAAACATACAGAACACAAAATGGGTCT
CACTTCTCTTCAAGTTTGATGGATTCTGATTGGCTCCAGGAATCCGAGTCATCAGGAGG
AAGCATGTTAGACTCTTCAACGAATTCTCCGT CAGCAGCCGACATACTAGCAGCTTGCAG
CACTAGACCACAAGCCTCGGCCGTGGCTGTAGCCGCTGCAGCTCTGATGGACGGTGAAG
GAGGCTGCGTCCACCTCAGGACCATCCTCAAAGTGTCTCGTTGCGAGTCAACACATAC
TAAGTTCTGTTACTACAATAACTACAGCCTCTCTCAGCCTCGTTACTTCTGCAAGACTTG
TCGCCGTTACTGGACAAAAGCGGGAACCTAAGGAATATTCCGGTTGGTGGTGGATGCCG
TAAAAACAAGAAACCATCTCTCTAATTCCTCCTCCTCCACTTCTTCCGGCAAAAAACC
ATCCAACATCGTTACCGCCAATACCTCTGATCTTATGGCTTTAGCACATTCTCATCAAAA
TTACCAACATTCTCCTCTAGGGTTTTACATTTTGGTGGGATGATGGGGTCTTACTCAAC
TCCGGAGCATGGTAACGTTGGTTTCTTGGAGAGCAAGTATGGCGGTTTGCTTTCGCAGAG
CCCTAGACCTATTGATTTCTTGGACAGTAAGTTTGATCTCATGGGAGTGAACAATGACAA
CCTGGTCATGGTTAATCATGGAAGTAACGGAGATCATCATCATCATATAATCATCACAT
GGGTCTGAATCACGGTGTAGGTCTTAACAACAACAACAATGGTGGATTTAATGGGAT
TTCTACGGGAGGCAATGGAAATGGTGGTGGTCTCATGGATATATCGACATGCCAAAGACT
TATGCTATCTAATTATGATCATCACCAATTACAATCATCAAGAAGATCATCAAGGGTAGC
AACAATAATGGATGTGAAGC AAAATCCGAAGTTGTTATCGCTTGATTGGCAGCAAGATCA
ATGCTACTCCAATGGTGGTGGTAGCGGAGGCGCAGGAAAATCCGACGGTGGTGGATACGG
CAATGGTGGTTATATCAACGGTTTAGGTTCTGTCGTGGAATGGTTTGATGAATGGCTATGG
AACGTCCACTAAAAEAACTCCTTGGTTTGATAAGTTAATCAGAACTCTTTTTTCTTGT
CGTCATCAACTAGTAGTAGTAGTAATAGTAGTTGGAGACTAGAGAAGCACTTCAAATTAT
TTATGGGTTTGTCTTGAAGCCAGTTTAC

>G375 Amino Acid Sequence (domain in AA coordinates: 75-103)

MGLTSLQVCMDSWLQSESSGGSM LDSSTNSPSAADILAACSTRPQASAVAVAAAALMD
GGRRLRPPHPDHPQKCPRCESHTKFCYNNYSLSQPRYFCKTCRRYWKGGTLRNIPVGG
GCRKNKKPSSNSSSSSTSSGKKPSNIVTANTS DLMALAHSHQNYQHSPLGFSHPGMMGS
YSTPEHGNVGFLESKYGGLLSQSPRPIDFLDSKFDLMGVNNDNLVMVNHGSNGDHHHHHN
HHMGLNHGVGLNNNNNGGFNGISTGGNGNGGGLMDISTCQRLMLSNDHHHHYHQEDHQ
RVATIMDVKPNPKLLSLDWQDDQCYSNNGGSGGAGKSDGGGYNGGYYINGLGSSWNGLMN

GYGTSTKTNSLV*

>G1007 (86..763)

ATTCCTTCTTGCCCTAGGAACATAATTGTTGCACACTTCGGTACACAATTTTTTGAGCACTT
CGACATCAAAACGAGAGAGAAAAGAATGGTGGATTCTCATGGCTCCGACACGGAATGTTT
CTCCAAGAAGAAAAAGGAGAAAAACGAAAGAAAAGGGGTATATCGTGGGGCTCGCATGAG
GAGCTGGGGGAAATGGGTCTCGGAGATTGGGGAGCCCCGTAAGAAATCAAGAATCTGGCT
CGGGACTTTCCACCGCGGAGATGGCAGCGCGTGCCCATGATGTTGCGGCATTGAGTAT
CAAAGGAAGTTCGCAATCCTTAACCTCCTGAGCTCGCGGATTTTCTGCCAAGACCACT
CTCGCTCAGCCAACAGGATATCCAGGCCGAGCCGCCGAAGCCGCTCTTATGGATTTCAA
AACTGTACCATTCATCTTCAGGATGACTCAACGCCGTTGCAAACTAGGTGTGATACTGA
GAAGATCGAAAAGTGGTCATCCTCATCGTCTCAGCCTCATCCTCATCCTCATCTTCGTC
CTCGTCTCATCTATGCTTTTCGGGGGAGCTAGGAGATATTGTGGAGTTGCCGAGTCT
TGAAAAAATGTAAATACGATTGTGCGCTGTATGACTCGTTGGAGGGGCTGGTGTGAT
GCCCCCATGGTTAGATGCTACCGAAAATGATTTTAGGTATGGAGATGATTCGGTACTGTT
GGACCCATGTCTCAAAGAAAGCTTTTGTGGAATTATGAGTAAGGTTTTTTTGGAAAG
AAATGTGGTTTTTTGTTCTCTCTCTTTTATACTTTTCGATCTTTTTTTCTAAGCATAT
ATATCTTCTACATATGTAATACTTTTCATTAGTAAACAATGATTCGGTTTCGGGTACAA
AAAAAAAAAAAAAAAAAAAAA

>G1007 Amino Acid Sequence (domain in AA coordinates: TBD)

MVDSHGSDETESSKKKKEKTKKEGVYRGARMRSWGKQVSEIREPRKKSRIWLGTFPTAEM
AARAHDAALSIKSSAILNFPBLADFLPRPVLSLQQDIQAAAAEALMDFKTVPFHLQD
DSTPLQTRCDTEKIEKWSSSSSSSSSSSSSSSSSSSSSMLSGELGDIVELPSLENNVKYDC
ALYDSLEGLVSMPPWLDATEFRYGDSDVLLDPLCKESFLWNYE*

>G1010 (344..1276)

ATTCCTTCTTCTAAAAAATCTTGACAACTTTTGTTTTGTTTTCTTTCTCTGAATTTTTT
AAAAGAGAGAGAGCTATGTAGCTATGAAACAGTAAGAGATATAGATATAGAGAGACAGAG
AAAGATGATGATCAGTGAAGTTAGGCTAAACCCACTTTCTATTTATGTATAATTAGGTCA
ATCACATACCAATCTCCTCCTCAATCTCCTCCTCCTCCTCCTCAAAATCTAGGGTTTTG
CTTGATCTCACCCCTTTCTCAATCCCTAGGGAAGCTGTAATTTTCATCAAAATCCAT
TATTTTTTGGTCACACCCCTTAAAGAGATCTGAGAGTTCTAAAGATGATGACAGATTTATC
TCTCAGGAGAGATGAAGATGAAGAAGCAAGCCCTTAGCAGAAGAAGAGAGCGCG
TGAAGTAGCAGACAGAGACATGTTTCGACAAAGTTGTGACTCCAAGTGATGTCGGAAA
ACTAAACCGACTTGTGATCCCAAAGCAACACGAGAGAGATTCTTCCCTTTAGATTATC
TTCAAACGAGAAAGGTTTGTCTTTTAACTTCGAAGATCTCACTGGCAAATCTTGGAGGTT
CCGTTACTCTTACTGGAACAGTAGTCAAAGCTATGTCATGACTAAAGTTGGAGCAGATT
CGTTAAAGACAAAAAGCTTGACGCCGAGATATTGTCTCTTTCCAAAGATGTGTGCGAGA
TTCAGGAAGAGATAGCCGTTTGTATTATTGATTGGAGGAGAAGACCTAAAGTCCCTGACCA
TCCTCATTTCCGCCCGGAGCTATGTTCCCTAGGTTTTACAGCTTTCCTTCGACCAATTA
CAGTCTTTATAATCATCAGCAGCAACGTCATCATCAGTGTTGGTGGTTATAATTATCA
TCAAATTCAGAGAGAAATTTGGTTATGGTTACTTCGTTAGGTCAAGTGGATCAGAGGAACAA
TCCTGCGGCTGCGGTGGCTGATCCGTTGGTGATTGAATCTGTGCCGGTGATGATGCACGG
GAGAGCTAATCAGGAACCTGTTGGAACGGCCGGAAGAGACTGAGGCTTTTTTGGAGTTGA
TATGGAATGCGGCGAGAGCGGAATGACCAACAGTACGAGGAGGAATCATCATCTTCCGG
TGGAAGTTTGCCACGTGGAGGCGGTGGTGGTCTTCATCTTCTCTTTTCAGCTGAG
ACTTGAAGCAGCAGTGAAGATGATCACTTCACTAAGAAAGGAAAGTCTTCATTGTCTTT
TGATTGGATCAATAATAATGATGATGATGAAATTAGTTGGTATTTTAAGAAAAAACA
TACATATATAAATCTATATATATGACAACATAATGCATTGATTTCTCT

>G1010 Amino Acid Sequence (domain in AA coordinates: 33-122)

MMTDLSLTRDEDEEEAKPLAEEEGAREVADREHMFQVVTTPSDVGKLNRLVIPKQHAERF
FPLDSSSNEKGLLLNFEDLTGKSWRFYSYWNSSQSYVMTKGWSRFVKDKKLDAGDIVSF
QRCVGDSDGRDRLFIDWRRRPKVPDHPHFAAGAMFPRFYSFPSTNYSLYNHQQQRHHHS
GGYNYHQIPREFGYGYFVRSVDQRNPNAAVADPLVIESVPVMMHGRANQELVGTAGKRL
RLFGVDMECGESGMTNSTEESSSSSGSLPRGGGGGASSSSFFQLRLGSSSEDDHFTKKG
KSSLSFDLDQ*

>G1014 (174..1112)

CACAAACCACAGTCTCTCTTCTCTCTATCTATCTTCTCTTCTCTCTCTATCTCTAT
CACTGAAACCCAAAGAGATCCACCATTTGTTCTTTTTCTCTTTCACACAGAGAAGTCTTT

CTTCCACACTTCCTTTTACTAGGCAGTGTTAACCAATTGAGAGAGAAAAATGATGGTTG
ATGAAAATGTGGAAACCAAGGCCTCTACTTTAGTGGCAAGTGTGATCATGGGTTTGGAT
CCGGGTCGGGTCATGATCATCATGGGTTATCGGCGTCTGTGCCTCTTCTTGGTGTAACT
GGAAGAAGAGAAGGATGCCTAGACAGAGACGATCTTCTTCTTCTTTAACCTTCTCTCTT
TCCCTCCTCTATGCTCCTATTTCCACGTGCCAATCCTCTCCCCGCACGTAAAATTG
ACCCAAGAAAGCTAAGATTCTCTTCCAAAAGGAAGTCAAGAACAGTGACGTGAGCTCTC
TCCGACGTATGATACTCCCGAAGAAAGCCGCGGAGGCTCACTTGCCGGCAGTTGAATGCA
AGGAAGGATTCCCTATAAGAAATGGAAGATTGACGGTTCACGTTTGGACCTTCAAGT
ATAGGTACTGGCCAAACAACAATAGCAGAATGTACGTGCTAGAAAACACAGGCGATTTTG
TGAATGCTCATGGTCTGCAGCTAGGTGACTTCATCATGGTTTACCAAGATCTCTACTCAA
ACAATTACGTTATACAAGCAAGAAAAGCATCGGAAGAAGAAGTAGACGTAATCAATC
TTGAAGAAGACGACGTTTACACAAACTTAACAAGGATCGAAAACACTGTGGTTAACGATC
TTCTCCTCCAAGATTTAATCATCACAACAACAACAACAACAGCAACAGCAACA
GCAACAAATGTTCTTACTATTATCCAGTCATAGATGATGTCACCACAAACACAGAGTCTT
TTGTCTACGACACGACGGCTCTTACCTCCAACGATACTCCTCTCGATTTTTTGGGTGGAC
ATACGACGACTACTAATAATTATTACTCCAAGTTCGGAACATTGATGGTTTGGGCTCCG
TTGAGAATATCTCTCTCGATGACTTCTACTAGATAATCAATCGATGGGCTCATGGTATTC
TTGATGGTGATCAGCTATTTAATATCCTTATAATATATAAGAATTAAATGCAATTTGC
ATATATATTATCAAGTGTGTGAATATAACATTACAGTTTTAAAAAAAAAAAAAAAAAAAA
>G1014 Amino Acid Sequence (domain in AA coordinates: 90-172)

MVDENVETKASTLVASVDHFGSGSGHDHGLSASVPLLGVNWKRRMPRQRSSSSFNL
LSFPPMPPIISHVPTPLPARKIDPRKLRFLFQKELKNSDVSLRRMILPKAAEAHLPAL
ECKEGIPIRMEDLDGFHVWTFKYRYWPNNNSRMYVLENTGDFVNAHGLQLGDFIMVYQDL
YSNNYVIQARKASEEEVDVINLEEDDVYTNLRIENTVVDLLQDFNHNNNNNNNSN
SNSNKCSYYPVIDDVTTNTESFVYDITALTSDNTPDLFLGGHTTTTNNYYSKFGTFDGL
GSVENISLDDFY*

>G1035 (103..624)

CCATAATAATATATAAACTATATACTATAATCTTTTTACATAATAAACTTTGGGTCTT
GCGTCTTAATCATAGTACTTAATTTCTCTGTGTGTTTAAATATGAATAATAAACTGAA
ATGGGATCTTCCACAAGTGGAATTGCTCGTTCGGTTTCAACCACTGGTTTAGCTAACTCC
GGTTCAGAATCTGATCTCCGGCAACGTGATCTAATCGACGAGCGGAAGAGAAAGAGGAAA
CAGTCGAACAGAGAATCTGCGAGGAGGTCGAGGATGAGGAAGCAGAAGCATTTGGATGAT
CTCACTGCTCAGGTGACTCATCTACGTAAAGAAAACGCTCAGATCGTCGCCGAATCGCC
GTCACGACGCAGCACTACGTCACTATCGAGGCGGAGAACGACATTCTCAGAGCTCAGGTT
CTTGAACCTTAACCACCGTCTCCAATCTCTTAACGAGATCGTTGATTTCTGCGAATCTTCT
TCTTCAGGATTCCGTATGGAGACCGGTGAGGATTAATTCGACGGTGGATTATTCGACGGC
GTGATGAATCCTATGAATCTAGGGTTTTATAATCAACCAATCATGGCTTCTGCTTCTACT
GCTGGTGATGTTTTCAACTGTTAGAAAACCTCACATCATTATCATCGTGAGTGAGACTAA
TCATCGCAGCAGGGGTAAAACGTAAATTTTTCTTATAAATTATGTGATGATGCTTTGTTT
CTTTATTTTATAAGATGGTTAATTAGTGTAAAACTGATTGTAATGATAGACAGTGTA
GAAATGTGTGATATCATGGAGATGGTGATGTGAGTTTGGTACAAATATTTAAGATCTTT
TCTTTCTATATATAAAGTGAAGAAATAATATTTGTCAATTTCTTAAAAAAAAAAAAAA
AAA

>G1035 Amino Acid Sequence (domain in AA coordinates: 39-91)
MNNKTEMGSSTSGNCSSVSTTGLANSSESIDLQRDLIDERKRKRKQSNRESARRSRMRK
QKHLDDLTAQVTHLRKENAQIVAGIAVTTQHYVTIEAENDILRAQVLELNHRLQSLNEIV
DFVESSSSGFGMETGQGLFDGGLFDGVMNPMNLGFYNQPIMASASTAGDVFNC*

>G1046 (1..567)

ATGATTAGACATCTAAAACCTACATGGAGTCGTCTAGTGTCATCGCTCTCATTTGTTTC
GATATTCTTGATGGAGTCCCACTACACGACGATCATTTCAACTCGGCATTCTACCAAAAC
ACTGACTTTAATGTTTCAATTTGCAGTCAAACGTATCGACCCGCATCAACAATCAGTCTCAC
TTAGACCCAAATGCAGAAAACATTTTCCATAACGAAGGTCTTGCTCCAGAAGAAAGAAGA
GCAAGAAGAATGGTCTCTAACCGGGAATCTGCAAGGAGGTACGTATGCGCAAAAAGAAG
CAGATCGAAGAGCTGCAACAACAAGTTGAACAACCTCATGATGTTGAATCATCACTTGTCT
GAGAAAGTCATCAACTTGTGGAAGCAACCATCAGATCCTACAAGAGAACTCACAGCTG
AAAGAGAAAGTCTCTTCTTCACTTGCTCATGGCAGATGTGCTATTACCCATGAGAAAT
GCAGAGAGCAACATCAATGACCGCAATGTGAATTATCTAAGAGGAGAACCATCAACCGT

CCCACCAACAGTCCCTTTGGTAAGTAA

>G1046 Amino Acid Sequence (conserved domain in AA coordinates:79-138)

MIRHLKPYMESSSVHRSHCFDILDGVPLHDDHFNSAFLPNTDFNVHLQSNVSTRINNQSH
LDPNAENIFHNEGLAPEERRARMVSNRESARRSRMRKKQIBELQQQVEQLMMLNHHLS
EKVINLLESNHQILQENSQLEKVKSSFHLLMADVLLPMRNAESNINDRNVNVLRGEPNSR
PTNSPFGK*

>G1049 (29..550)

CTAACTTTCTTCCCAAGTAACTTCAAAATGCAGCCGCAAACAGACGTTTTCAGCCTCCA
TAACTACCTAAACTCATCGATACTGCAGTCTCCGTATCCTTCTAATTTCCCGATATCTAC
GCCATTTCCAACCAACGGTCAAAACCCGTACCTCCTCTACGGATTCAAAGCCCTACAAA
CAATCCACAATCCATGAGCCTAAGCAGCAACAACCTCAACATCAGATGAAGCAGAAGAGCA
GCAGACGAACAACAATAATCAACGAGCGGAAGCAGAGAAGGATGATTTCAAACCGAGA
ATCCGCAAGGAGATCGCGTATGAGGAAGCAAAGACACCTTGACGAGCTTTGGTCACAAGT
GATGTGGTTAAGGATCGAGAATCATCAGTTGCTTGATAAGCTTAACAATCTCTGAGTC
TCACGACAAGGTTCTTCAAGAGAATGCTCAGCTTAAAGAAGAACATTTGAGCTTAAGCA
AGTGATCAGCGATATGCAAATCAAAGCCCTTTCTCTTGCTTTAGAGACGATATAATCCC
CATTGAATAAAGCATTTTTCCCGATTATATTTATGAAAATTTCTTCAAGAGTATGTT
TCTTTGTATGTATATGTGGAGATGATTTTCAGGGTTTGTATAATATGACCCTTACGACG
ACGTTTTTAGATTGTAGTAAATTTATAAACTAAAGAAGATTAGTGTTAATGAAGAACAAA
TATAA

>G1049 Amino Acid Sequence (domain in AA coordinates 77-132)

MQPQTDVFLSHNYLNSSILQSPYPNFIPTFPPTNGQNPYLLYGFQSPNTNPNQSMSSLS
NNSTSDAEAEQQTNNNIINERKQRRMISNRESARRSRMRKQRLHDELWSQVMWLRIENHQ
LLDKLNNLSESHDKVLQENALKEETFEKQVISDMQIQSPFSCFRDDIPIE*

>G1069 (89..934)

TTGGAACCTAGAGGCTTTCAAGCAAATCATCAGGGTAACAATTTCTTGATCTTTCTTT
TTAGCGAATTTCCAGTTTGTGTCATCATGGCAAACCTTGGTGGACGAACCAGAGTGG
TTTAGCGGGCATGGTGGACCATTCGGTCTCCTCAGGCCATCACCAAAACCATCACCACCA
AAGTCTTCTTACCAAAGGAGATCTTGAATAGCCATGAATCAGAGCCAAGACAACGACCA
AGACGAAGAAGATGATCTTAGAGAAGGAGCCGTTGAGGTGGTCAACCGTAGACCAAGAGG
TAGACCACCAAGGATCCAAAAACAAACCCAAAGCTCCAATCTTTGTGACAAGAGACAGCCC
CAACGCACTCCGTAGCCATGTCTTGGAGATCTCCGACGGCAGTGACGTGCGCGACACAAT
CGCTCACTTCTCAAGACGCAAGCAACGCGCGTTTGCCTTCTCAGCGGGACAGGCTCAGT
CGCTAACGTCAACCTCCGCCAAGCCGCGCACAGGAGGTGTGGTCTCTCTCAAGGCAG
GTTTGAAATCTTATCTTTAACCGGTGCTTTCTCCTCGGACCTTCCCCACCCGGGTCAAC
CGGTTTAAACGGTTTACTTAGCCGGGGTCCAGGGTCAGGTGCGTTGGAGGTAGCGTTGTAGG
CCCACTCTTAGCCATAGGTCGGTTCATGGTGATTGCTGCTACTTTCTCTAACGCTACTTA
TGAGAGATTGCCCATGGAAGAAGAGGAAGACGGTGGCGGCTCAAGACAGATTACCGGAGG
CGGTGACTCACCGCCAGAATCGGTAGTAACCTGCCTGATCTATCAGGGATGGCCGGGCC
AGGCTACAATATGCGCGCCGATCTGATTCCAAATGGGGCTGGTCAGCTAGGGCACGAACC
ATATACATGGGTCCACGCAAGACCACCTTACTGACTCAGTGAGCCATTTCTATATATAAT
GGTCTATATAAATAAATATATAGATGAATATAAGCAAGCAATTTGAGGTAGTCTATTACA
AAGCTTTTGTCTGCTGGTTGGAAAAATAAATAAGTATCAAAGCTTTGTTTGTCTTAAATGGA
AATATAGAGCTTGGGAAGGTAGAAAGAGACGACATT

>G1069 Amino Acid Sequence (domain in AA coordinates: 67-74)

MANPWWTNQSLAGMVDHSVSSGHHQNHQSLLTGDLGIAMNQSQDNDQDEEDDPREG
AVEVNNRRPRGRPPGSKNPKAPIFVTRDSPNALRSHVLEISDGSVDVADTIAHFSRRRQR
GVCVLSGTGSVANVTLRQAAPGGVVSQGRFEILSLTGAFPLPGSPPGSTGLTVYLAGV
QQQVVGGSVVGPLLAIGSVMVIAATFSNATYERLPMEEEDGGGSRQIHGGDSPPRIGS
NLPDLSGMAGPGYNMPPHLIPNGAGQLGHEPYTWVHARPPY*

>G1070 (170..1144)

TCGACCAGCTTGGATTTCTGTTTCATCATTACTACTCTCTTTCTTCTCTAGCTAGCTA
GTTTTGACAGCAAAATAAGAAAGCAAAAAAAGGTCAACTAAAAAAGATCTGTTCTTAGAT
CACTCTCTTCTTTCTTTTGTATCCAATTCCACCATTGAATCATAGATCATGGATCCAGT
ACAACTCTCATGGATCACAAAGCTCTCTACCTCCTCCTTTCCACGCAAGAGACTTTCAATT
ACATCTTCAACAACAGCAACAAGAGTTCTTCTCCTCCACCATCACCAGCAACAAGAAACCA
AACCAGTGGTGACCAACAAGGAGGATCAGGAGGAAACCGACAAATCAAGATGGATCGTGA

AGAGACAAGCGACAACATAGACAACATAGCTAACAACAGCGGTAGTGAAGGTAAAGACAT
AGATATACACGGTGGTTTCAGGAGAAGGAGGTGGTGGCTCCGGAGGAGATCATCAGATGAC
AAGAAGACCAAGAGGAAGACCAGCGGGATCCAAGAACAACCAAACACCGATTATCAT
CACACGGGACAGCGCAACCGCGCTTAGAACCCACGTGATGGAGATCGGAGATGGCTGCGA
CTTAGTCGAAAGCGTTGCCACTTTTGCACGAAGACGCCAACCGCGCGTTTGCCTTATGAG
CGGTACTGGAAATGTTACTAACGTCATATACGTACGCTGGATCTCATCCTTCTCCTGG
CTCGGTAGTTAGTCTTACCGGAAGGTTTCGAGATTCTATCTCTCTCAGGATCTTTTCTCCC
TCCTCCGGCTCCTCTACAGCCACCGGATTGAGTGTTTACCTCGCTGGAGGACAAGGACA
GGTGGTTGGAGGAAGCGTAGTTGGTCCGTTGTTATGTGCTGGTCTGTGCTTGTATGGC
TGCGTCTTTTAGCAATGCGGCGTACGAAAGGTTGCCTTTAGAGGAAGATGAGATGCAGAC
GCCGGTTCATGGCGGAGGAGGAGGAGGATCATTGGAGTCCGCCCAATGATGGGACAACA
ACTGCAACATCAGCAACAAGCTATGTCAAGTCATCAAGGGTTACCACCTAATCTTCTTGG
TTCGGTTAGTTGCAGCAGCAACATGATCAGTCTTATTTGGTCAACGGGACGACCACCGTA
TTGATCAATATACACACACTCATAATCGTTGCTAGCTAGCTAACGATGAATCATGAG
TTTAGTGGATATATATATGATTAAAGAGGTTAGCTTATGAACATTAATAAGAGTTTGA
TTCTATCGAGCTTCATTATGTTTGGGTCATCGTTC

>G1070 Amino Acid Sequence (domain in AA coordinates: 98-120)
MDPVQSHGSQSSLPFPHARDPQLHLQQQQQEFFLHHHQQQRNQTGDQGGSGGNRQIK
MDREETSNDNIDNIANNSSGSEKIDIDIHGSGSEGGGSGGDHQMTRRPRGRPAGSKNPKP
PIIITRDSANALRTHVMEIGDGC DLVESVATFARRRQRGVCMVSGTGNVTNVTIRQPGSH
PSPGVSVSLHGRFEILSLSGSFLPPPAPPTATGLSVYLAGGQGVVGGSVVGPLLCAGPV
VVMAASFNAAYERLPLEEDEMOTPVHGGGGGSGLESPPMMGQQLQHQQQAMSGHQGLPP
NLLGSVQLQQQHDQSYWSTGRPPY*

>G1076 (198..1076)
ATTTTAGTCTTCTATAACTTCTTCTCAATCCTCTCTCATATCTTTTTTCTTAGTTTAAA
TTTCAATAAAATAGAAAAAACATATACAAATCTACAGAGAAGAGAAGCTTTATTTTAAT
CTTGTGTGTGTGTGTGTGTGTTTATATAATTTTTATTTTTCCTTCAATTAATAATCTCTCT
TTGCTTTTGATGTGGGATCGGCTGGTCTTGATCTAGGCACAGCTTTTCGTTACGTTAATC
ACCAGCTCCATCGTCCGATCTCCACCTTCACCACAATTCTCTCCTCCGATGACGTCACCTC
CCGGAGCCGGGATGGGTCAATTTACCGTCCGACGACGAGACAACAACAACCATCAAG
GTCTTGACTTAGCCTCTGGTGGAGGATCAGGAAGCTCTGGAGGAGGAGGATCACGGCG
GGGGAGGAGACGTCGTTGGTCTGTCGTCACGTCGACGACGAGGAGGAGGATCCAAGAACAAC
CGAAACCTCCGGTAATTATCAGCGCGAGAGCGCAAACTCTAAGAGCTCACATTCTTG
AAGTAACAAACGGCTGCGATGTTTTGACTGCGTTGCGACTTATGCTCGTCGGAGACAGC
GAGGGATCTGCGTTCTGAGCGGTAGCGGAACGGTCACGAACGTCAGCATACGTCAGCCAT
CTGCGGCTGGAGCGGTTGTGACGCTACAAGGAACGTTCCGAGATTCTTCTCTCTCCGGAT
CGTTTCTTCTCTCTCCGGCACCTCCCGAGCAACGAGTTTGACAATTTCTTAGCCGGAG
GACAAGGTGAGGTGGTTGGAGGAAGCGTTGTGGGTGAGCTTACGCGGCTGGACCGGTGA
TTGTGATTGCAGCTTCGTTTACTAATGTTGCTTATGAGAGACTTCCTTTAGAAGAAGATG
AGCAGCAGCAACAGCTTGAGGAGGATCTAACGGCGGAGGTAATTTGTTTCCGGAGGTGG
CAGCTGGAGGAGGAGGAGGACTTCCGTTCTTTAATTTACCGATGAATATGCAACCAAATG
TGCAACTTCCGGTGGAAGGTTGGCCGGGAATTCGGGTGGAAGAGGTCTTTCTGATGTG
TATATATTGATAATCATTATATATATACCGGCGGAGAAGCTTTTCCGGCGAAGAATTGTC
GAGAGTGAAGAAAGGTTAGAAAAGCTTTTAATGGACTAATGAATTTCAAATTATCATCGT
GATTTCCGGACATTGTCTTGTTCATCATGTTAAGCTTAGGTTTATTTTGTGCTTGTAG
AATTTTATGTTGAATCTTTTTTTTTTCTGTGAACTCTATTGTGTTCTGCTGCGAAGG
AAAAAAAATTCTCAAAAAAA

>G1076 Amino Acid Sequence (domain in AA coordinates: 82-89)
MAGLDLGTAFRYVNHQLHRPDLHLHNSSSDDVTPGAGMGHFTVDEDDNNNNHQLDLAS
GGGSGSSGGGGHGGGDDVGRPRGRPPGSKNPKPPVITRESANTLRAHILEVTNGC
DVFDCAVYARRRQRGICVLSSGTVTNVSIHQPSAAGAVVTLQGTFFELSLSGSFLPPP
APPGATSLTIFLAGGQGVVGGSVVVELTAAGPVIVIAASFNTVAYERLPLEEDEQQQQL
GGGSNGGGLNLFPEVAAGGGGLPFFNLPMNMQPNVQLPVEGWPGNSGGRGPF*

>G1089 (31..2427)
AAGTAAGAGAGCTTCTTAAGGAAGAAGAAGATGGGTTGTGCTCAATCAAAGATCGAGAAC
GAAGAAGCAGTTACTCGTTGCAAAGAACGAAAACAATTGATGAAAGACGCCGTCAGTGCT
CGTAACGCTTTCGCCGCCGCTCACTCAGCTTACGCTATGGCTCTTAAAAACACCGGAGCT

MGCAQSKIENEEAVTRCKERKQLMKDAVTARNAFAAAHSAAYAMALKNTGAALSDYSHGEF
LVSNNHSSSSAAAAIASTSSLPTAISPLPSSSTAPVSNSTASSSSAAVPQPIPDTLPPPPP
PPPLPLQRAATPEMNGRSGGGHAGSGLNGIEEDGALDNDDDDDDDDDSEMNDRDLIR
KRSRSGGSTRGNRTTIEDHHLQEEKAPPPPLANSRPIPPPRQHQQHQQQQQQFFDYDF
FPNVENMPGTTLDEFPPOPQOPTRPVPPQPHSPVITTEDDEEEEEEEEEEEETVIER
KPLVEENPKRVEEVTELEKVTNLGRMKKSQIGVIGERRGMRMPVTATHLANVIELND
NFLKASESAHDVSKMBEATRLHYHSNFADNRGHIHDSARVMRVITWNRSFRGINADDDGK
DDVDLEENETHATVLDKLLAWEEKLYDEVKAGELMKIEYQKKVAHLNRVKKRGGHSDSLE
RAKAAVSHLHTRYIVDMQSMDSSTVSEINRLRDEQLYLKLVLHLEAMGKMWEMMQIHHQRQ
AEISKVLRLSDVSAQVETNDHHHERTIQLLAVVQVWHTQFCRMDHQKEYIKALGGWLK
LNLPIESTLKEKVSSPPRPVNPAPIQKLLHAWYDRLDKIPDEMAKSAIINFAAVVSTIMQ
QQEDEISLRNKCEETRKEKLRKIRQFEDWYHKYIQKRGPEGMPNPDEADNDHNDDEVAVRQF
NVEQIKRLEEBEEAYHRQSHQVREKSLASLRLPELFQAMSEVAYSCSMDMYRAITYAS
KROSOSERHOKPSOGHYS*

>G1093 (1..531)

ATGGGTTATCCGGTGGGTACACTGAGCTCCTCCTCCCAAGAATCTTCCTTCACTTACTC
TCTCTCTTAGGCTTAATACGAACACTCATAGACACGGGTTTTTCGGATATTGGGTCTACCC
GACTTTCTCGAATCCGACCCGGTTTCATCGTCATCGTCATGGCTGGAACACCGTATATG
TCCACGGCGGCGCATCATCACCAGAAAGCTCATTTTTCTTCCCAGTGGCGGCGAGGCTA
GCTGGAGAAATCTTGCCCGTCATCAGATTCTCGGAGCTAACTCGACCCGGATTCCGATCC
GGATCCGATTGCTGCGCGGTGTGCCTCCACGAGTTCGAGAACGATGACGAGATCCGACGG
CTGACGAATTGTCAACACATATTTACCCGGAGCTGTTTAGACCGTTGGATGATGGGTAT
AATCAGATGACGTGTCCACTTTGTAGAACGCCGTTTATTTCTGATGAGTTACAAGTTGCT
TTTAACCAACGAGTTTGGTCTGAATCTGAACCTTCTCGCAGAATCAAATTAG

>G1093 Amino Acid Sequence (domain in AA coordinates: 105-148)

MGYPVGYTELLLPRIFLHLLSLGLIRTLIDTGFRILGLPDFLESPPVSSSSSWLEPPYM
STAAHHHQESSFFFFPVAARLAGEILPVIRFSELTRPGFSGSDCCAVCLHEFENDDEIRR
LTNCQHIFHRSCLDRLMMGYNQMTCPLCRTPFISDELQVAFNRVWSESELLAESN*

>G1127 (191..1351)

GACAGACTCTCTGTATGTGTGCGAGAAGCGAGAAGCGAGAGAGAGAGAGAGAGAGTTG
TTAGCTCACACGCTTTCTCTATTTTCTCGGAATTCACAAAACAGAAAGTTTCATCCTTTA
CGAGAATTAAAGCCGAAAGAAACAATCTTTGAGTTTGATTCTTCTTCTTCTTCTTCTCT
CTCTGCTCTAATGGATTCCAGAGACATCCACCGTCACATAACCAGCTTCAACCACCACC
GGGAATGTTAATGTCTCATTACCGTAACCTAACGCCCGCTTACCATTAAATGGTTCC
CACTTCCACATCTCAACCGATTCAACACCCTCGTCTTCTTTTGGCAATCAACAACAATC
TCAAACGTTTCATCAGCAGCAACAACAACAATGGATCAGAAGACTCTTGAATCTCTTGG
ATTTGGTGATGGATCACC'TTCTTCTCAACCGATGCGATTCCGGATCGATGATCAGAATCA
GCAACTGCAAGTGAAGAAGAGCGAGGAAGGCCGAGAAAGTATACCTCTGATGGTAGCAT
TGCTTTAGGTTTAGCTCCTACGTCTCTTCTTCTTCTGTCAGCTTCTAATTCTTACGGTGA
GGGTGGTGTGTGAGATAGTGGTGGAAATGGAACTCTGTTGATCCACCTGTTAAACGTAA
CAGAGGAAGGCTCCTGGTTCTAGTAAGAAACAGCTTGATGCTTTAGGAGGAACCTCAGG
AGTTGGGTTTACACCTCATGTCAATTGAAGTGAACACAGGAGAGGACATAGCGTCAAAGGT
GATGGCTTTTTCGGATCAAGGGTCAAGAACAATTTGTATTCTCTCTGCAAGTGGTGCAGT
TTCTAGAGTGATGCTCTCGTCAAGCTTCTCATTCTAGTGAATCGTTACTTATGAGGGACG
ATTTGAGATCATTACTCTCTCAGGCTCAGTCTTGAATTATGAGGTAAATGGTTCCACCAA
CAGAAGTGGTAACTTGAGTGTGGCTTTGGCTGGACCTGATGGCGGCATCGTAGGTGGCAG
TGTAGTTGGTAACTTAGTAGCTGCAACACAAGTCCAGGTGATAGTGGGAAGCTTTGTTGC
AGAAGCAAAGAAACCGAAACAAGTAGTGTTAACATTGCTCGGGGGCAGAATCCTGAACC
GGCTTCAGCGCCGGCTAACATGTTGAAC'TTGGATCAGTCTCTCAAGGACCATCGAGCGA
GTCATCAGAAGAGAATGAGAGCGGTTCTCCTGCAATGCACCGTGACAATAATAATGGGAT
ATATGGAGCTCAACAACAACAACAACAACCTCTTCATCCTCATCAGATGCAAATGTA
CCAACATCTTTGGTCTAATCATGGTCAATAAAATGAAGCGGAAATTAATTTGTTTCCGTT
TTGGTTACGGTTATGGTTTGATTCTT

>G1127 Amino Acid Sequence (domain in AA coordinates:103-110, 155-162)

MDSRDIPPSHNQLQPPPGMLMSHYRNPNAASPLMVPSTSQPIQHPRLPFGNQQSQT
HQQQQQQMDQKLTLESFGDGSPSSQPMRFGIDDQNLQVKKRGRPRKYTPDGSIALG
LAPTSPLLSAASNSYEGGVGSDSGNGNSVDPPVKRNRGRPPGSSKKQLDALGGTSGVGF
TPHVIEVNTGEDIASKVMASDQGSRTICILSASGAVSRVMLRQASHSSGIVTYEGRFEI
ITLGSVNLNVEVNGSTNRSNLSVALAGPDGGIVGGSVVGNLVAATQVQVIVGSFVAEAK
KPKQSSVNIARGQNPASAPANMLNFGSVSQGPSSSESSEENESGSPAMHRDNNNGIYGA
QQQQQQQLPHPHQMOMYQHLWSNHGQ*

>G1131 (57..758)

TCGACTCCTCTCCTGATTGCTTCACCTTCTTCTTTACTACAGGTTTCAGCTCCTCAATGT
CCATGGATTGCTTAAGCTACTTCTTTAACTACGATCCTCCTGTCCAGCTCCAGGATTGCT
TTATTCCCGAGATGGATGATTATCCCTGAAACCGATAGTTTCTTCTTCCAATCTAAC
CGCAACTGGAGTTTCATCAGCCATTGTTCAAGAAGAAGCTCCTTCACAGACCCACTTTG
ACCCTTTCTGCGACCACTTTCTTCTCCGCAAGAAATCTTCTCCCTAACCTAAAAACG
AAATCTTCAACGAAACACACGACCTCGATTCTTCTTCTCCCAACGCAACGCGAGAGAC
TTGTTAACTCCAGCTACAATTGTAACTCAAAACCATTTCCAGAGCCGTAACCCGAATT
TCTTCGACCCCTTTCGGCGACACTGATTTCGTTCCAGAATCTTGTACCTTCCAGGAGTTTC
GAGTTCGGGATTCTCTTTAGCTTTCAAGGTAGGCCGGGGAGATCAAGATGACTCAAAGA

AACCGACGCTTTCATCTCAGAGCATCGCGGCTAGAGGGAGGAGAAGAAGAATTGCAGAGA
AGACTCACGAGCTCGGAAAACTCATCCCCGGTGGCAATAAACTTAACACCGCCGAGATGT
TCCAAGCCGCGCTAAGTATGTCAAGTTTTTGCAGAGTCAAGTTGGGATTCTCCAAGTGA
TGCAGACCACAAAGAAGGTAATAACCAACCCCAAATAAGAAGTTTATCATCCAATTGAAA
CTCTAATCGTGTCTTCTCACAAGCTTCTTAATTTGTTTACGCAGGGTAGCTCTAATGTGC
AAATGGAACTCAGTATTTGCTTGAATCGCAAGCAATCCAGGAGAAGTTATCAACAGAGG
AAGTGTGTTTGGTACCGTGTGAAATGGTTCAAGATCTAACAAGTGAAGAAACCATTTGCA
GAACCCCGAATATTTCTCGAGAAATCAACAAGTTACTGTCTAAACATCTGGCTAACTAGT
TTTAGTTTCAAGCCTGAAGTTCTCTATGCCTAAATTTGTGTCTGTTATCGTTGTTTGTCT
TCTTAGTTAGTGTGTTTGTCTTGTGATTAGGGGCTAATTATCCTGGTTAATCTCCTCT
TAACTGGGAA

>G1131 Amino Acid Sequence (domain in AA coordinates: 173-220)
MSMDCLSYFFNYDPPVQLQDCFIPEMDMIIPETDSFFFQSQPQLEFHQPLFQEEAPSQTH
FDPFCDQFLSPQEIFLPNPKNEIFNETHDLDFFLPTPKRQRLVNSSYNCTQNHQSRNP
NFFDPFGDTDFVPESCTFQEFVRPDLFLAFKVGKRGDQDDSKKPTLSSQSIARGRRRRIA
EKTHELGKLIIPGKNLNTAEMFQAAAKYVKFLQSQVGILQLMQTTKKVITNPK*

>G1145 (243..1142)
GTGATTTCTCTCTGCCATTTCTTCGATTGATTCTGGGTTCTCTTCTCTCTCTCTCTC
TTCTGCATGTTTTCGCCACTCTACCTTAGAAAAAGGTTACTTTTCGCCTCCGATTTAGGCT
CGATTTGATGAATTCGTCGTCGTGTGGCTATTTATCAAATTGAGCATTAGGGTTTCTGAT
TTGTGGGTTTCAGAATGTTTTATCTATCTGTCTTGTGTTTTTGTCCGCTACAAAAGC
CTATGGATTCTCAGAGGGGTATTGTTGAACAAGCTAAATCTCAGTCTTGAATAGGCAAA
GCTCTCTTTACAGCTTAAACATTGATGAGGTTCAAAATCACTTGGGGAGTTCTGGTAAAG
CTCTGGGAAGCATGAACCTTGATGAGCTTTTGAAGAGTGTCTGTTCTGTTGAAGCTAATC
AGCCATCGCTGTATGGCTGTCAATGGTGGAGCAGCTGCTCAGGAGGGTCTTTCTCGCCAGG
GGAGTTTGACTTTGCCTCGGGATCTCAGCAAAAAGACTGTTGATGAGGTTTGGAAGACA
TTCAGCAGAATAAGAATGGAGGTAGTGTCTCATGAGAGGAGGGATAAGCAGCCTACACTTG
GGGAAATGACGCTTGAAGACCTGTTGTTGAAAGCAGGAGTGGTCACTGAGACTATCCCTG
GTTTCAACCATGATGGTCTGTTGGTGGTGGTAGTCTGGTTTCAAGTGTCTGGTTTAGGGC
AAAACATTACTCAAGTTGGCCCATGGATTCAATATCATCAGCTCCCATCAATGCCACAGC
CTCAAGCATTTATGCCCTATCCGGTTTTCAGATATGCAAGCAATGGTGTCTCAGTCTTCTT
TGATGGGTGTTTGTCTGATACACAAAACCTCTGGAAGGAAGAGGGTAGCTTCAGGAGAAG
TTGTAGAGAAGACTGTAGAGAGGAGGCAGAAGAGAATGATAAAGAACAGAGAGTCTGCTG
CTCGTTCCCGAGCTAGGAAAAGGCTTACACTCATGAGCTAGAGATCAAAGTTTACGGT
TAGAAGAAGAAAACGAAAGACTCAGGAAGCAAAAGGAGGTGGAAGAAATCCTCCCAAGTGT
ACCACCGCTGATCCCAAGCGGCAGCTCCGACGGACAAGCTCGGCTCCTTTCTGATCTCT
AAACTCTTTTGTCTTTTCTTTTTTCTTCTGTGTCTGTTCACTTATAAAAAAGAGA
GGAAAACAGCTTTGTTTCTTTGTACATTCCGTAGACTTTCTTGACTTGGAGCAATTCTGT
TAACCTTTAAATATTCTCGAGTTATTGTAGTAGCAGACTAGCAGCAGTAATGGTTTTCAT
GAGTCCGATTGAAATTCAGAGATTGAACAGGAAAAAA

>G1145 Amino Acid Sequence (conserved domain in AA coordinates: 227-270)
MDSQRGIVEQAKSQSLNRQSSLYSLTLDEVQNLGSSGKALGSMNLDELLKSVCSVEANQ
PSSMAVNGGAAAQEGLSRQGSLLTLPRLDSKKTVDVWKDIQONKNGGSAHERRDKQPTLG
EMTLEDLLLKAGVVTETIPGSNHDGPVGGGSAGSGAGLQONITQVGPWIQYHQLPSMPQP
QAFMPYPVSDMQAMVSQSSLMGGLSDTQTPGRKRVASGEVVEKTVERRQKRMKNRESAA
RSRARKQAYTHELEIKVSRLEENERLRKQKEVEKSSQVYHRLIPSGSSDQARLLSDL*

>G1229 (123..1217)
TTTGGGCGGGTCTTTCTTTCCCTAAATCTTTCTTTTATTTTGCTGTTTAAAAAAAATC
CAACCATAAGACAAAACAACGAACGAGGAAGAGAGAGAGAAGGATATATCTCTAATCA
CGATGCAGGAGATAATACCGGATTTTCTTGAAGAGTGTGAATTTGTCTGACACTTCACTAG
CCGGAGATGATCTATTTGCCATCTTAGAGAGTCTTGAAGGTGCGGAGAGATATCTCCGA
CAGCTGCATCTACACCTAAAGATGGAACCAAGTTCCAAGGAGTTAGTTAAGGATCAAG
ATTATGAAAACCTCATCTCTTAAAGAGGAAAAAGCAAGACTAGAAACCAGGAAAGAAGAGG
ACGAAGAAGAAGAAGACGGAGACGGAGAAGCAGAAGAAGATAATAAGCAAGATGGGCAAC
AAAAGATGTCTCATGTAACCGTGAACGTAACCGGAGAAAGCAAATGAACGAGCACTTAA
CCGTTTTGCGTTCTCTTATGCCTTGTCTTCTACGTCAAACGGGGGACCAAGCATCGATCA
TAGGAGGAGTTGTGGAGTACATAAGCGAGTTACAACAAGTTCTCCAATCTTTGGAAGCCA

AGAAACAACGTAACCTACGCCGAAGTCCTAAGCCCGAGAGTTGTCCCGAGCCCTCGTC
CTTCACCGCCTGTTCTAAGCCCAAGAAACCGCCTCTTAGCCCGCGCATCAACCACCACC
AGATTACCACCACCTACTTCTCCCTCCCATAGTCCCTCGAACACCTCAGCCAACAAGCC
CATACCGGGCCATTCCACCGCAACTACCACTCATCCACAGCCTCCGCTTCGCTCTTACA
GCTCATTTGGCCAGTTGTCAGCAGCTTAGGAGATCCACCTCCATACTCTCTGCTTCATCTT
CTTCATCTCCTTCAGTTAGTAGTAACCATGAGAGTAGTGTGATCAATGAGCTTGTGCTA
ACTCAAAATCGGCTTTGGCTGATGTGGAAGTGAAGTTTTCAGGAGCTAACGTGCTGCTCA
AAACGGTGTGTCGATAAGATCCCGGGACAAGTTATGAAGATAATTGCTGCTCTTGAAGATT
TGGCTCTTGAGATTCTTCAGGTTAATATTAACACCGTCGACGAAACCATGCTTAATTCTT
TCACCATCAAGATTGGAATTGAGTGCCAACTAAGTGCAGAAGAACTGGCTCAACAAATTC
AGCAAACATTCTGCTAGTAAAGAAGGATTAAATATAGCTTCGTATAAACCTTAACGAGAG
AGCAGTACGTACTACTTCTCTCTCTTAGTATCCCTTTAATTATCTTTTCAGTTTTCTGC
AAAGATATGGAGTTTAAAAAATAAAATTTGTTATCTAAAGTTTAAATCAAATATTGATTA
ATTATAACTAATATAGGTATAAGTGAGTTTAAAGATTATCAGCTTCATAACAGCCATCG
TCATGTTTACTTTCTTTTAAATTTTAGAATTTAGACGTACTCCTACCATGTAATTTTATT
TCTGTCTATACATCAAGCATTGTAGCTGTAATTGCATATGAATGAACAATAGTGTATGAG
TGATCTCATGAATAATATTCTTCTTGCAACACAAAAA

>G1229 Amino Acid Sequence (domain in AA coordinates: 102-160)

MQEILPDFLEEEFVDTSLAGDDLFAILESLEGAGEISPTAASTPKDGTSSKELVKDQD
YENSSPKRKKQRLERKEEDEEEDGDGEAEEDNKQDQQKMSHVTVERNRRKQMNHLT
VLRSLMPCFYVVRGDDQASIGGVVEYISELQVQLQSLQLEAKKQKRTYAEVLSPRVVPSRP
SPVLSPRKPLSPRINHHQIHHLHLLPPISPRTPQPTSPYRAIPPQLPLIPQPLRSYS
SLASCSSLGDPPPYSPASSSSPSVSSNHSSVINELVANSKALADVEVKFSGANVLLK
TVSHKIPGQVMKIIALEDLLEILQVNINTVDETMNSFTIKIGIECQLSAEELAQQIQ
QTFC*

>G1246 (1..1746)

ATGATCATGTACGGAGGAGGAGGAGCAGGGAAGGACGGTGGATCCACCAATCACTTATCA
GACGGAGGAGTGATATTGAAGAAAGGTCCATGGACGGCGCGGAAGATGAGATACTTGCT
GCGTACGTTAGAGAGAACGGTGAAGGGAATTGGAACGCCGTTTCAGAAAAACACAGGTTT
GCTCGTTGCGGCAAAAGCTGCCGCTCTCGATGGGCCAATCACCTCCGACCAAATCTGAAA
AAAGGCTCTTTCACCGGTGACGAAGACGTCTCATCATTCAGCTTCATGCTCAGCTTGGT
AACAAATGGGCTCGCATGGCTGCTCAGTTACCGGGAAGAACAGACAACGAGATTAAAGAAC
TATTGGAACACGAGATTGAAACGACTTCTTCGCCAAGGACTTCCTCTTTATCCTCCAGAT
ATTATCCCTAACCATCAACTCCATCCACATCCACATCATCAACAACAACAGCAACATAAC
CATCATCATCATCATCAACAACAACAACATCAACAATGTATTTTCAACCACAA
TCTTCACAACGAACACACCATCATCTTCCCCTCTTCCATCTCCAACACCAGCAAAACGCA
AAGTCTCATCATCTTCACTTTTCATACCACGACTGCTAACCTCTCCATCCACTTAGC
CCTCACACTCCAAACACACCATCTCAACTCTCTTCCACACCGCCTCCACCACCACTTTCC
TCTCCTTTATGTTCCCTTCGCAACAACCAATACCCGACCTTCCCCTCTTTGCCCTCCCG
CGTTCCCAAAATCAACAACAACAACCGGAAATTTCACTTTCCCTAGACCTCCACCTCTC
CTTCAACCGCCTTCATCACTCTTCGCAAAACGTTACAACAATGCTAACACTCCTCTTAAT
TGCATCAACCGCGTCTCAACCGCACCATTTTCCCCTGTTTCAAGAGACTCCTACACTTCC
TTTCTTACATTGCCCTTACCCTTCCCCAACCGCTCAAACCGCTACTTACCACAATACTAAT
AACCCTTACTCTTCTCTCTCTCTCTCTTTAAACCTTCTTCTTCTTCTTACCCTACA
TCAACTTCTTCCCCAAGCTTTCTTCACTCCCATTAACACTCCTTCTTCCACCTCATTTTAT
ACCAACCCAGTTTACTCCATGAAACAAGAGCAGTCCCTTCAAACCAAATCCCCAAATA
GATGGCTTCAATAACGTCAACAACCTTCAAGACAACGAGAGACAGAATCATAACCTTAAC
AGTTCCGGTGTCTAFAGAAGAAGTAGTAGCTGCAGCCTCTTAGAGGATGTCTTCAAGAG
GCCGAAGCTTTAGCTCTGAGGCGAGGCGGACCTCCAAAACGAAGACAACCTCACAGCT
TCTCTTCCGAACCAACAACAACCAACAACAACGACAACCTTCTTCTCGGTTAGTTTC
GGACATTATGATTCTTCTGACAACCTATGTTTCTTGAAGATTGAAATCAAAGGAAGAA
GAGTCTCTTCAAATGAACACAATGCAGGAGGACATAGCTAAGCTTCTTGATTGGGGAAGT
GATAGTGGAGAGATCTTAATGGACAATCATCTGTGTCACTGACGACAATCTGTTCTT
GATGTTTCATCAATTAGCTTCACTATTTCCCGGTGATTCTACAGCCGTCTAGCCGCAACA
AACGACCAACAACAAGAATAATAACAATAATTGTTCTGCGATGACATGCAGGGAATA
AGGTAG

>G1246 Amino Acid Sequence (domain in AA coordinates: 27-139)

MIMYGGGAGKDDGGSTNHLSDGGVILKKGPTAAEDEILAAYVRENGEGNWNNAVQKNTGL
ARCGKSCRLRWANHLRPNLKKGSFTGDEERLIIQLHAQLGNKWARMQAQLPGRTDNEIKN
YWNTRLKRLRLRQGLPLYPPDIIPNHQLHPPHQQQQQHNNHHHHHHQQQQQHQQMYFQPPQ
SSQRNTPSSSPLPSPTPANAKSSSFTFHTTTANLLHPLSPHTPNTPSQLSSSTPPPPPLS
SPLCSPRNNQYPTLPLFALPRSQINNNNGNFTFPRPPPLLPSSSLFAKRYNNANTPLN
CINRVSTAPFSPVSRDSYTSFLTLPYPSPTAQATYHNTNNPYSSSPSFLNPSSSSYPT
STSSPSFLSHYTPSSSTFHTNPVYSMKQEQLPSNQIPQIDGFNNVNNFTDNERQNHNLN
SSGAHRRSSSCSLEDVFEEAEALASGGRGRPPKRRQLTASLPNHNNTNNNDNFFSVSF
GHYDSSDNLCSLQDLKSKEESLQNMNTQEDIAKLLDWGSDSGEISNGQSSVVTDDNLVL
DVHQLASLFPADSTAVVAATNDQHNKNNNNNCSWDDMQGIR*

>G1255 (138..1388)

CAGCTCAAACCTCTCTAGGACTACACTAAATCTAACTTTTTGCAGAGAGCAAAAGATTCAA
TAATTGAGATTGATCTCAAAACCAAGCTCTCGTGTCTTGTCTGTGATGTTGGTTGTGT
AGACTTTGTATACAATGATGAAAAGTTTGGCGAATGCTGTTGGAGCGAAGACGGCGAGGG
CTTGCGACAGCTGCGTGAAGAGACGTGCACGGTGGTACTGCGCGGCCGACGATGCTTTTC
TTTGCCAGTCTTGCGACAGTTTGGTCCATTACAGCAAACCTCTTGTCTGCGCCGCCACGAGA
GAGTCCGTTTGAAGACGGCTAGCCCGGCGGTCTGTAAGCATAGCAACCACTCATCAGCTT
CTCCTCCACATGAGGTGCGCCAGTGGCATCACGGGTTTACTCGTAAAGCTCGAACGCCAC
GTGGCTCTGGTAAAGAAAACAATTCTGTCGATATTTTCATGACTTGGTTCCTGATATTAGTA
TTGAGGATCAGACAGACAACTATGAGCTTGAAGAGCAGCTGATCTGTCAAGTGCCGGTTC
TAGATCCGTTGGTGTCTGAGCAGTTCTGAACGATGTCGTTGAGCCCAAGATCGAGTTTC
CTATGATCAGAAGTGGTTTGTATGATCGAGGAGGAGGAAGACAACGCTGAAAGTTGTCTTA
ATGGATTTTCCCGACCGACATGGAGCTTGAGGAGTTTGTCTGCTGACGTGGAGACTCTGC
TCGGTCCGCGGTTAGACACGGAGTCGTATGCCATGGAGGAGCTAGGGTTATCTAATTCA
AGATGTTCAAAAATCGAAAAGATGAGATTGAAGAAGAACTAGAAGAGATAAAAGCCATGA
GCATGGATATATTTGATGATGATCGAAAAGACGTGGATGGAACAGTACCGTTTGAGCTAA
GCTTTGATTACGAGTCGTACACAAGACGTCCGAAGAAGAGGTAATGAAGAACGTTGAAA
GTAGTGGTGAATGTGTTGTTAAGGTGAAAGAGGAAGAACATAAGAATGTTCTGATGCTAA
GATTAACTATGACTCGGTGATATCCACTTGGGGAGGTCAAGGTCCACCGTGGAGTTCAG
GAGAGCCACCGGAACGAGACATGGACATCAGCGGTTGGCCAGCCTTTTCCATGGTGGAGA
ATGGAGGAGAAAGTACTCATCAGAAGCAATACGTTGGTGGATGTTTACCATCAAGTGGGT
TTGGAGATGGAGGTAGAGAAGCTAGAGTTTCGAGATACAGAGAGAAGAGGAGGACAAGGT
TGTTTTCTAAGAAGATACGGTACGAGGTACGTAAATTGAATGCAGAGAAAAGACCACGAA
TGAAAGGAAGATTCTGTAAGAGAGCCTCGCTCGCTGCTGCTTACCATTAGGTGTTA
ATTACTGAATAGTTAATATCTATTATCTCACTTTACAAATTTCCGTGAATCT
TTTTTCTTCTGAAACAACAGAAGTTATTTTGGCACTTAATGTGCTTTGAGGACTTGTAT
GTACATAGAAGTAACCAATAATAATGTGACTTTTACTA

>G1255 Amino Acid Sequence (domain in aa coordinates: 18-56)

MKSLANAVGAKTARACDSCVKRRARWYCAADDAFLCQSCDSLVSANPLARRHERVRLKT
ASPAVVKHSNHSASPPHEVATWHHGFTRKARTPRGSGKKNSSIFHDLVPDISIEDQTD
NYELEEQLICQVPVLDPLVSEQFLNDVVEPKIEFPMIRSGLMIEEEEDNAESCLNGFFPT
DMELEEFADVETLLGRGLDTESYAMEELGLSNSEMFKIEKDEIEEEVEBIKAMSMDIFD
DDRKDVDGTVPFELSFDYESHKTSEEEVMKNVSSGECVVKVKEEHNKVLMLRLNYDS
VISTWGGQGPWPSSGEPPEPDMDISGWPAPFSMVENGGESTHQKQYVGGCLPSSGFGDGG
EARVSRVREKRRLTFSKKIRYEVRLNLAERPRMKGRFVKRASLAAAASPLGVNY*

>G1304 (1..978)

ATGGGGCGATCACCATTGTTGCGATGAGAATGGTCTAAAGAAAGGGCCATGGACACAAGAG
GAGGATGATAAACTGATAGATCACATTCAAAAACATGGCCATGGCAGCTGGAGAGCTCTT
CCAAAGCAAGCCGGTTTAAACCGATGCGGAAAGAGTTGTAGATTAAAGATGGACCAACTAC
TTGAGACCTGACATCAAGAGAGGAAATTTCACTGAAGAGGAAGAACAACTATTATCAAC
CTCCATTCCCTTCTTGGAACAAGTGGTTCGTCGATAGCCGGTAATCTTCTGGAAGAACG
GACAATGAAATAAAAACTATTGGAACACACATTTGAGAAAGAACTTCTCCAAATGGGG
ATTGATCCGGTGACCCATAGGCAAGAACCGACCATCTAAACGTTTTAGCAGCTCTCCCG
CAGCTTATAGCCCGCCGCAAAATTTCAACAGCCTCTTGAATCTCAACCAAAATGTGCAACTG
GATGCAACAACCTTGTCTAAAGCTCAACTGCTACACACTATGATTCAAGTCCTTAGCACC
AATAACAACACCACCAATCCTTCTTTTCTTCACTCAACTATGCAAAACAGTAACACCAAT
CTCTTTGGCCAAGCTTCTTACTTAGAGAACCAAAATCTTTTGGTCAGTCTCAAACTTC

TCTCACATTCTTGAGGATGAGAATTTGATGGTCAAACCCAAATTATTGATAACCCCTTG
GACTCTTTTCTTCCCCCATACAACCCGGTTTTCAAGATGATCATAATTCACCTCCCTCTA
TTGGTTCGCGCTCTCCTGAAGAATCTAAAGAACTCAAAGGATGATCAAGAACAAGAC
ATCGTCGATTACCATCATCATGATGCTTCAAACCCCTCATCATCAAACCAACGTTTACA
CAAGATCATCATCACCCATGGTGTGACACTATTGATGATGGAGCAAGTGATTCTTTTGG
AAAGAGATAATAGAGTAA

>G1304 Amino Acid Sequence (conserved domain in AA coordinates:13-118)

MGRSPCCDENGLKKGPTWQEEEDKLIDHIQKHGHSWRALPKQAGLNRCGKSCRLRWNY
LRPDIKRGNFTEEEQTIINLHSLGKNWSSIAGNLPGRTDNEIKNYWNTHLRKKLLQMG
IDPVTHRPRTDHLNVLAAALPQLIAANFNSSLNQLNQNVLQDATTAKAQLLHTMIQVLST
NNNTTNPSFSSSTMQNSNTNLFQASYLENQNLFGQSQNFHILEDENLMVKTQIIDNPL
DSFSSPIQPGFQDDHNSLPPLVPASPEESKETQRMINKNDIVDYHHHDASNPSSNSTFT
QDHHHPWCDTIDDGASDSFWKEIIE*

>G1318 (7..849)

AAAAATATGAGGAAGCCAGAGGTAGCCATTGCAGCTAGTACTACCAAGTAAAGAAGATG
AAGAAGGGACTTTGGTCTCTCGAGGAAGACTCAAAGCTGATGCAATACATGTTAAGCAAT
GGACAAGGATGTTGGAGTGATGTTGCGAAAAACGCAGGACTTCAAAGATGTGGCAAAAGC
TGCCGTCTTCGTTGGATCAACTATCTTCGTCTCTGACCTCAAGCGTGGCGCTTCTCTCCT
CAAGAAGAGGATCTCATCATTCGTTTTCATTCCTCCTCGGCAACAGGTGGTCTCAGATT
GCAGCACGATTGCCTGGTTCGGACCGATAACGAGATCAAGAATTTCTGGAACCAACAATA
AAGAAAAGGCTAAAGAAGATGTCCGATACCTCCAACCTAATCAACAACCTCATCCTCATCA
CCCAACACAGCAAGCGATTCTCTTCTAATTCGCGATCTTCTTTGGATATTAAAGACATT
ATAGGAAGCTTCATGTCTTACAAGAACAAGGCTTCGTCACCCCTTCCTTGACCCACATA
CAAACCAACATCCATTTCCAACGGGAAACATGATCAGCCACCCGTGCAATGACGATTTT
ACCCCTTATGTAGATGGTATCTATGGAGTAAACGCAGGGGTACAAGGGGAACCTTACTTC
CCACCTTTGGAATGTGAAGAAGGTGATTGGTACAATGCAATATAAACAACCACTTAGAC
GAGTTGAACACTAATGGATCCGGAACGCACCTGAGGGTATGAGACCAGTGGAAGAATTT
TGGGACCTTGACCACTGATGAACACTGAGGTTCCTTCGTTTTACTTCAACTTCAAACAA
AGCATATGAATATTTTACGTATCTTATCTTTTTTCTATTGCGGTTTATACTCAAGAT
TCTTAGCCACACACATAAATGCAAATATATATACATTGTTAGAGAGTATTTTGTATTT
CGTATAATCTTTTCGTACTAGGGCTTGAGCCTTGAGGTCCCATGTAACGATTAGTCAATG
TAAACATATATCTATAATAAATAAAATAAAAGAAATAATAAGCACATAAAAAAAAAAA
A

>G1318 Amino Acid Sequence (domain in AA coordinates: 20-123)

MRKPEVAIAASTHQVKMKKGLWSPEEDSKLMQYMLNNGQGCWSDVAKNAGLQRCGKSCR
LRWINYLRLPDLKRGAFSPQEEEDLIIRFHSILGNRWSQLAARLPGRTDNEIKNFWNSTIKK
RLKMSDTSNLIINNSSSSPNTASDSSSNSASSLDIKDIIGSFMSLQEQGFVNPSLTHIQ
NNPFTGNMISHPCNDDFTFYVDGIYGVNAGVQGEYFPFPLECEEGDWNANINNHDEL
NTNGSGNAPEGMRPVEFWDLQLMNTEVPSFYFNFKQSI*

>G1320 (39..788)

GAAGATCATAAAGATCAAAGGAGAGAGGTATTAATAAATGATGTGTAGTCGAGGCCATT
GGAGACCTGCAGAAGACGAGAAGCTAAGAGAACTCGTCGAGCAATTTGGTCCTCATAATT
GGAACGCCATAGCTCAGAAGCTCTCTGGTTCGATCTGGTAAGAGTTGTAGATTGAGATGGT
TTAATCAATTGGATCCTAGGATTAACCGAAACCTTTACGGAGGAAGAAGAAGAAAGGC
TTTTAGCGCTCATCGGATCCATGGGAACAGATGGTCTGTGATCGCTAGATTTTTTCCCG
GTGAACTGATAACGCTGTATAAAACCATTTGGCACGTCATCATGGCTCGTCGTGGCCGAG
AACGGTCCAAGCTCCGTCCACGAGGCCCTTGGCCATGATGGCACGGTGGCTGCGACTGGGA
TGATTGGTAATTATAAAGACTGCGATAAGGAGAGAAGATTGGCAACCACAACCGCTATCA
ATTTTCCTTATCAATTCTCTCATATTATCATTTTTCAAGTCCTCAAAGAGTCCTTGACCG
GAAAGATCGGGTTCAGAAATAGTACTACTCCAATACAAGAAGGAGCAATAGACCAAATA
AACGACCGATGGAGTTCTACAATTTCTCCAAGTAAACACGGATTGGAAGATACACGAAT
TGATAGATAATTCAAGAAAAGACGAAGAAGAAGATGTCGATCAAAACAACCGAATTCGTA
ACGAGAATTGTGTTCCATTTTTCGACTTTTTGTCTGTGGAAACTCTGCCTCTCAGGGTT
TATGTTAATTTGTCCGTACCATGTACTATAAGGTGGACCATATGTTAACTAAAGATAA
TGTAAGAAAGTACTAATCAATTAGAGCTCTGTTTGAGCCAAATGTGAAAATTAGTTAAGA
CATCCCAACATTTTCTGTATAACACATATAAGGTTGTACTTTTATCAGGTCTAATTTT
CTATTTTATTTTAAGGATGTTTAAATCAGACCCATAACCATTCGATAAAAAAAAAAAAAA

>G1320 Amino Acid Sequence (domain in AA coordinates: 5-108)
MMCSRGRHWRPAEDEKLRELVEQFGPHNWNIAIAQKLSGRSGKSCRLRWFNQLDPRINRNP
TEEEEEERLLAPHRIHGNRWSVIARFFPGRTDNAVKNHWHVIMARRGRERSKLRPRGLGHD
GTVAATGMIGNYKCDKERRLATTTAINFPYQFISHINHFQVLKESLTGKIGFRNSTTPIQ
EGAIDQTKRPMFYNFLQVNTDSKIHLEIDNSRKDEEEDVDQNNRIRNENCVPFFDFLSV
GNSASQGLC*

>G1330 (36..959)

GTACCGGCGACCTCTTTGTGGGTCACTCTTCATCAATGGGTGACAAAGGAAGGAGCTTAA
AGATCAACAAGAACATGGAGGAATTCACGAAAGTGGAGAAGAAATGGACGTAAGGAGAG
GTCCATGGACAGTTGAGGAAGATTTAGAGCTCATCAATTACATTGCTAGTCATGGTGAAG
GTCGATGGAACTCTCTCGCTCGTTGCCGGAACCTCAAAGGACCGGAAAAAGCTGCAGAC
TTCGGTGGCTGAACTATCTCCGACCAGATGTGCGCCGTGGAAACATAACCTCGAAGAAC
AACTCTTGATTCTTGAACCTCACACACGTGGGGCAATAGATGGTCTAAGATTGCACAAT
ATTTACCAGGAAGAACGATAACGAGATCAAAAACCTATTGGAGAACACGTGTTCAAAGC
ATGCAAAACAGCTTAAATGCGACGTGAACAGTCAACAATTTAAAGACACCATGAAGTATC
TTTGGATGCGCTCGGCTCGTAGAAAGGATCCAAGCCGCGTCCATCGGGTCTGTTTCCATGT
CATCTTGCGTACCACCTCTCAGATCAGTTCGTGATCAACAACAACACCAACAACG
TGGATAATTTGGCTTTTAAATGAGTAACCTAATGGTTACATCACGCCGGAATAATTCCAGCG
TGGCAGTATCTCCTGTATCAGATTTGACGGAGTGTCAAGTGAGTAGTGAAGTGTGGAAGA
TTGGTCAGGATGAGAATTTGGTGGATCCAAAATGACATCGCCGAATTATATGGATAATA
GCAGTGGACTATTAAACGGAGATTTTACGAAGATGCAAGATCAAAGTGACCTTAATTGGT
TTGAAAATATTAAATGGGATGGTACCAATTTATTCGGACAGTTTTTGAACATTGGAAATG
ATGAAGACTTCTGGCTCTTACAACAACATCAACAAGTCCACGACAATGGAAGCTTCTGAA
TAGACAAGAAGCTATGCGGCC

>G1330 Amino Acid Sequence (domain in AA coordinates: 28-134)
MGDKGRSLKINKNMEEFKVEEEMDVRRGPWTVEEDLELINYIASHGGRWNSLARCAEL
KRTGKSCRLRWLNLYRPDVRGNITLLEQLLILELHTRWGNRWSKIAQYLPGRDNEIKN
YWRTRVQKHAKQLKCDVNSQQFKDTMKYLWMPRLVERIQAASIGSVSMSSCVTTSSDQFV
INNNNTNNVDNLALMSNPNGYITPDNSSVAVSPVSDLTECQVSSEVWKIGQDENLVDPKM
TSPNYMDNSSGLNGDFTKMQDQSDLNWFENINGMPVNYSDSFWNIGNDEDFWLLQHQHQ
VHDNGSF*

>G1352 (79..900)

GCGCGATTAAAACTCTCAACTTTTCTCTCAAATTTCTGATCCTTTGATCCAACAGTTAG
AAGAAGATTTCATCTGATCATGGCCCTCGAAGCGATGAACACTCCAACCTCTTCTTTTACC
AGAATCGAAACGAAAGAAGATTTGATGAACGACGCCGTTTTCATTGAGCCGTGGCTTAAA
CGCAAACGCTCCAAACGTGAGCGTTCTCACAGCCCTTCTCGTCTTCTTCTCACCAGCT
CGATCTCGACCCAAATCCAGAATCAAGATCTTACGGAAGAAGAGTATCTCGCTCTTTGT
CTCCTCATGCTCGCTAAAGATCAACCGTCGCAAACGCGATTTCATCAACAGTCGCAATCG
TTAACGCCGCCGCCGAGATCAAAGAACCTTCCGTACAAGTGTAACGCTGTGTAAGAAAGCG
TTTCTCTCTATCAGGCTTTAGGCGGTACAAAGCAAGTCACCGAATCAAACCACCAACC
GTAATCTCAACAACCGCCGATGATTCAACAGCTCCGACCATCTCCATCGTCGCCGGAGAA
AAACATCCGATTGCTGCCTCCGGAAAGATCCACGAGTGTCAATCTGTATAAAGTGT
CCGACGGGTCAAGCTTTAGGCGGTACAAACGTTGTCACTACGAAGGCAACCTCGGCGGC
GGAGGAGGAGGAGGAAGCAAATCAATCAGTCACAGTGGAGCGTGTGCGAGCAGGTATCG
GAAGAAAGGAGCCACCGTGGATTATCGATCTAAACCTACCGCGTTACCTGAACTCAGC
CTTCATCACAAKPTTIVISTTADDSTAPTISIVAGEKHPIAASGKIHECSICHKVFPTGQAL
GGHKRCHYEGNLGGGGGGGSKSISHSGSVSSTVSEERSHRGFIDLNLPALPELSLHNP
VDEEILSPLTGKKPLLLTDHDQVIKKEDLSLKI*

>G1352 Amino Acid Sequence (domain in AA coordinates: 108-129,167-188)

MALEAMNTPTSSFTRIETKEDLMNDVFIPLWLRKRKRQRSHSPSSSSSSPPSRPKS
QNQDLTEEEYLALCLLMLAKDQPSQTRFHQQSQSLTPPPESKNLPYKCNVCEKAFPSYQA
LGGHKASHRIKPTTIVISTTADDSTAPTISIVAGEKHPIAASGKIHECSICHKVFPTGQAL
GGHKRCHYEGNLGGGGGGGSKSISHSGSVSSTVSEERSHRGFIDLNLPALPELSLHNP
VDEEILSPLTGKKPLLLTDHDQVIKKEDLSLKI*

>G1354 (1..1047)

ATGGAAAGTCTCGCACACATTCCTCCCGTTATCGATTCCATCCGACCGATGAAGAACTC

GTTGACTATTATCTCAAGAACAAAGTTGCATTCCCGGGAATGCAAGTTGATGTTATCAAA
GATGTTGATCTCTACAAAATCGAGCCATGGGACATCCAAGAGTTATGTGGAAGAGGGACA
GGAGAAGAGAGGGAATGGTATTCTTTTAGCCACAAGGACAAGAAATATCCAAGTGGGACA
CGAACCAATAGAGCAACGGGCTCCGGATTGGAAGCAACGGGTCGAGACAAGGCCATT
TACTCAAAGCAAGAGCTTGTGGGATGAGGAAGACTCTTGTCTTTTACAAAGGTAGGGCC
CCAAATGGTCAGAAATCTGATTGGATAATGCACGAATACCGTCTTGAGACCGATGAAAAT
GGACCGCCTCATGAGGAAGGATGGGTGGTTGTGCGCGTTTCAAGAAGAAGCTAACCACG
ATGAACTACAACATCCAAGAACAATGATGGGATCATCATCAGGCCAAGAATCTAAGTGG
TTCACGCAGCAAATGGATGTGGGGAATGGTAATTACTATCATCTTCTGATCTAGAGAGT
CCGAGAATGTTTCAAGGCTCATCATCATCACTATCATCATTACATCAGAATGATCAA
GACCCCTTATGGTGTGCTACTCAGCACTATTAACGCAACCCCAACTACAATAATGCAACGA
GATGATGGTCATGTGATTACCAATGATGATGATCATATGATCATGATGAACACAAGTACT
GGTATCATCATCAATCAGGATTACTAGTCAATGATGATCATAATGATCAAGTAATGGAT
TGGCAAACGCTTGACAAGTTTGTGCTTCTCAGCTAATCATGAGCCAAGAAGAGGAAGAA
GTTAACAAGATCCATCAGATAATTCTTCGAATGAAACATTTTCATCATCTCTCTGAAGAG
CAAGCTGCAACAATGGTTTCGATGAATGCTTCTTCTCTCTCTCCATGTTCTCTCTAC
TCTTGGGCTCAAAATACACACAGTAA

>G1354 Amino Acid Sequence (domain in AA coordinates: TBD)
MESLAHIPPGYRFHPTDEELVDYLLKNKVAFFPMQVDVIKVDLYKIEPWDIQELCGRGT
GEEREWYFFSHKDKKYPTGTRNTRATGSGFWKATGRDKAIYSKQELVGMRLTLVFKGRA
PNGQKSDWIMHEVRLTDENGPPHEEGWVVCRAFKKLTMNYYNNPRTMMGSSSGQESNW
FTQQMDVGNNGNYHLPLDLESPRMFQSSSSSLSLHQNDDQPYGVVLSTINATPTTIMQR
DDGHVITNDDHMMIMNTSTGDHHSGLLVNDDHNDQVMDWQTLDFVASQLIMSQEEEE
VNKDPSDNSSNETFHLSEEQAATMVSMNASSSSSPCSFYSWAQNTHT*

>G1360 (1..1257)

ATGGGAGATAGAAACAACGACGGTGATCAGAAAATGGAGGATGTATTGTTGCCCGGATT
AGGTTTCATCCAACCGACGAAGAGCTCGTAAGCTTCTACCTGAAGCGGAAGGTTCAACAC
AACCTCTCTCCATGAGCTCATAAGACAACCTCGATATCTACAAATATGACCCCTGGGAT
CTTCCAAAGTTTGCGATGACGGGTGAAAAAGAATGGTACTTTTATTGTCCAAGGGACAGG
AAGTATAGGAACAGCTCGAGGCCAAACCGAGTGACCGGAGCTGGTTTTTGGAAAGCCACG
GGAACGGACCGGCCGATATACTCGTCAGAAGGAAACAATGCATAGGTTTAAAGAAGTCC
TTAGTGTTCTACAAAGGAAGAGCAGCGAAAGGAGTTAAGACTGATTGGATGATGCATGAG
TTTCGTTTGCCTTCTCTCTCCGAACCATCTCTCTCTTAAGAGATTCTTCGACTCTCCT
GTCTCTCCCAACGATTCATGGGCTATATGCAGAATCTTCAAAAAGACCAACACAACGACC
CTAAGAGCTCTCTCTCACTCTTTTGTTCCTCGTTACCACCAGAAACAAGCACCGACACA
ATGTCTAACCAAAAGCAATCAACACATACCATTTTTCTTCAGACAAGATCCTCAAACCT
AGCTCTCACTTCCAGTTTACCATGAGAATATGAACACTCCCAAACTAGTAATAGTACA
ACTCATCCGTTCCCACTATAAGTCCCTTCTCTTACTTGGATTTCACCTTCATACGACAAA
CCCACCAACGTTTTCAATCCGGTTTCATGTTTACACCAACAATACCTCACAAATCTCTTT
CTTGCCACACAAGAAACACAACCTCAGTTTCCAGGCTCCCTCGTCAAATGAAATCCCA
TCGTTTCTGCTAAACACGTCTTCAGATTGACCTTCTTGGGAGAATTCACGAGCCATATC
GACCTCAGCGCAGTGTTGGCCCCAAGAGCAATGTCCCCCGCTTGTAAGCCTACCACAGGAG
TATCAAGAGACGGGATTGCAAGGAAATGGTATAATGAAGAACATGCGTGGTTCCAATGAA
GATCATCTTGGTGATCATTGCGACACACTTCGGTTTGATGATTTCACTTCAACAATTAAT
GAGAACCATCGTCATCATCAAGACCTGAAACAGAACATGACATTGCTGGAGAGTTATTAT
TCTTCTTTATCGTCCATCAATAGCGATTGTCAGCTTGTCTTCTCCAGTACAACTGA

>G1360 Amino Acid Sequence (conserved domain in AA coordinates:18-174)

MGDRNNDGDQKMEDVLLPGFRFHPTDEELVSFYLLKRVQHNPLSIELIRQLDIYKDPWD
LPKFAMTGEKEWYFYCPRDRKYRNSRPNRVTGAGFWKATGTDRIYSSSEGNKICIGLKS
LVFYKGRAAKGVKTDWMMHEFRPLSLSESPSPSKRFFDSPVSPNDSWAICRIFKKTNTTT
LRALSHSFVSSLPETSTDMSNQKQSNTHYFSSDKILKPSSHQFHHENMNTPKTSNST
TPSVPTISPFSTLDFTSYDKPTNVFNPVSLDQQLTLNLFATQETQPPQFRLPSSNEIP
SFLNLTSSDSTFLGEFTSHIDLAVLAQEQCPPLVSLPQEQETGFEGNGIMKNMRGSNE
DHLGDHCDTLRFDDFTSTINENHRHHQDLKQNMTLLESYSSSLSSINSDLPACFSSTT*

>G1364 (1..537)

ATGGCGGAGTCGACGGCCAAGAGTCCCGGAGGCTGTGGAAGCCATGAGAGTGGTGGAGAT
CAAAGTCCAGGTCGTTACATGTTCTGAGCAAGATAGGTTTCTTCCGATTGCTAACATA

AGCCGTATCATGAAAAGAGGTCTTCTGCTAATGGGAAAATCGCTAAAGATGCTAAGGAG
ATTGTGCAGGAATGTGTCTCTGAATTCATCAGTTTCGTCAACGAGCGAAGCGAGTGATAAA
TGTCAAAGAGAGAAAAGGAAGACTATTAATGGAGATGATTGCTTTGGGCAATGGCTACT
TTAGGATTTGAAGACTACATGGAACTCTCAAGGTTTACCTGATGAGATATAGAGAGGGT
GACACAAAGGGATCAGCAAAAGGTGGGGATCCAAATGCAAAGAAAGATGGGCAATCAAGC
CAAAATGGCCAGTTCTCGCAGCTTGCTCACCAAGGTCTTATGGGAACCTCTCAAGTAAC
TTTCTCTCTCTCTCTTCACTCAAGCAATACGCATCATTCTCTTCTAATTTGTAA
>G1364 Amino Acid Sequence (conserved domain in AA coordinates:29-120)
MAESQAKSPGGCGSHESGGDQSPRSLHVREQDRFLPIANISRIMKRGLPANGKIAKDAKE
IVQECVSEFISFVTSEASDKCQREKRKTINGDDLLWAMATLGFEDYMEPLKVYLMRYREG
DTKGSAGKGDPPNAKDGQSSQNGQFSQLAHQGPYGNSSQVTFPLFSSSHSSNTHHSLIC*
>G1379 (68..622)
CTCTGCCTCTCTCTCTCTCTCAAAACCCATCTCGAAAGTCTTCTCTTTTCGAGGGTTTAG
ATCCTCCATGGAAGCGCGGAGTTGCTGACGTGGCTGTCCCGGTACGAGGAAGAGAGA
CAGACCTTACAAAGGAATTAGGATGAGGAAGTGGGAAAGTGGGTGGCGGAGATTCTGTGA
GCCTAACAGCGCTCTAGGTTATGGCTTGGCTCTTACTCTACTCCCGAGGCGGCGCGCG
AGCTTACGACACGGCGGTTTTCTATCTTAGAGGACCTACGGCGAGGCTTAACCTCCCTGA
GCTTCTTCTGAGGAGAAATTTCTCCGACGAGGATATGTCGGCTGCGACCATCAGGAAGAA
AGCCACGGAGGTGGTGTCTCAGGTTGATGCTTTGGGCACGGCGGTGCAAAATAACCGCCA
CCGTGTTTTTGGTCAGAATCGAGATACTGATGTGGATAATAAGAATTTTCATCGGAATTA
TCAAACGGTGAACGAGAAGAAGAAGAAGATGAGGATGACAAGAGATTGAGGAGTGG
CGGCCGGTTATTGGATCGGGTTGACTTGAATAAATTACCCGACCCGGAAGCTCCGATGA
AGAATGGGAAAGCAACATTAATAATATATAGTTTGGAGCGGTGGCTGTTGCTAACGTAC
GCCAACGGCTTGTCTTACGAATCATTAGCGCGGTTATGATTTTTTTTTTTTTTTTTTT
CATTATCTGAAAAATTTAGGGCTTTTTAGTTATTAATTTTGTGTTTGTGTTTTCTTTCT
TGCGAGTTTTCGGTTTATGGAATTTTAGGCTATTGCTTAACGAAAAAAAAAAAAAAAAA
>G1379 Amino Acid Sequence (domain in AA coordinates: 18-85)
MEGGGVADVAVPGTRKRDRPYKGIIRMKWGWVAEIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTARLNFPELLPGEKFSDEDMSAATIRKKATEVGAQVDALGTAVQNNRHRV
FGQNRSDVDNKNFHRNYQNGEREEEEDEDDKRLRSRGLLDRLVDLNLKLPDPESSEEW
ESKH*
>G1384 (33..977)
GTACATTTTTTTTTTGTATTTTCAGGAACTCCGATGGCGGATCTCTTCGGTGGTGGCCACG
GCGGCGAGCTTATGGAAGCACTTCAACCTTTTTACAAAAGTGCTTCCACGTCTGCTTCAA
ATCCTGCGTTTGCCTCCTCAAACGATGCGTTTGCCTTGCCTCCAAACGACCTATTTCTT
CTTCTTCTTACTATAATCCTCATGCATCTTATTCCCTTACATTCACAAACCTCTTACC
CGGATATTTATTCTGGATCCATGACCTATCCATCTTCAATTCGGGTTCGGATCTTCAACAAC
CCGAAACTACCAATCTCAGTTCCATTACCAAAACACTACCTTACACTCACCAGACA
ACAACTTCTTATGCTTAACTTCAATTGAGCCGAGCCAACCGGGTTTTATGACCCAACCGG
GTCCGAGTTCGGGTTTCGGTTTCAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGGAAATGGGTTCGGGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTTCGACACGGCTGAAGAAGCCGCTTGGCTTATGATCGCGCCGCTTTAAGCTTC
GTGGTGACTCGGCTCGGCTTAACTTCCAGCTCTCCGATACCAAACCGGCTCGTCTCCGT
CTGATACCGGCGAATATGGTCTTATTCAAGCTGCCGTAGACGCTAAACTAGAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCCGGCTTCTTACGCTGAGCAGCCGTCAGCGCCACAACAACATTCCGGGTTCGGGTGAAA
GTGATGGGTTCGGGTTACCGACTTTCGGATGTTATGGTGCAGGAGATGTGCCAAGAGCCAG
AGATGCCATGGAATGAAAATTTTCATGCTCGGCAAGTGTCTTCTTATGAGATAGATTGGG
CTTCAATTTTATCGTGAAAAATTAGGATTCAATTCATTTTATTCAATTTAACTTGTTTG
TATTTCTTTTAACTTTAGGGTTATTAGCTGTGCGTAA
>G1384 Amino Acid Sequence (domain in AA coordinates: TBD)
MADLFGGGHGGELMEALQPFYKSASTSASNPFASSNDFAFASPNDLFSSSSYNPHASL
FPSSHSTTSYPIYSGSMTYPSFSGSDLQQPENYQSQFHYQNTITYTHQDNNTCMLNFI
EQPGFMTQPGPSSGSVSKPAKLYRGVRRHWGKWVAEIRLPNRNTRLWLGTFTDAEEAAL
AYDRAAFKLRGDSARLNFALRYQTGSSPSDTGEYGPIQAAVDKLEAILAEPKNQPGKT
ERTSRKRAKAAASSAEQPSAPQQHSGSGESDGSPTSVMVQEMCQEPMPWNNFMLG
KCPSYEIDWASILS*

>G1399 (261..1475)

AGGTCGAATTTTCTGAAATTAAGATTCATTCTCCATGGAAGAAGCTCTGTTTTATTCT
CTTTAGCTTAGCTTAGCTTCTACTGATCTGTTTTTGCTACAAAATCCCATCTTTTCTTT
AAAACCTCTTATCTCTGAATCTTGAGTTTCTTGTAAGAAGAAGCAATTTTGAATCTTT
CGTAATCATAAAGATTCGTGGAGGATCTCTACTGATTGTGCGGAATCTCTCACTACAGAA
TCACCTTGATCTTATGTCCGGATGGAGGAGAGAGAAGGAACCAACATCAACAACACATCA
CTAGCAGTTTCGGCTTGAAGCAGCAACATGAAGCTGCTGCTTCTGATGGTGGTTACTCAA
TGGACCCACCACCAAGACCCGAAACCCCTAACCCGTTTTAGTCCCACCCACTACTGTCC
CCGCGGCCGCCACCGTAGCAGCAGCTGTTACTGAGAATGCGGCTACTCCGTTTAGCTTAA
CAATGCCGACGGAGAACACTTCAGCTGAGCAGCTGAAAAAGAAGAGAGGTAGGCCGAGAA
AGTATAATCCCGATGGGACTCTTGTCTGACTTTATCGCCGATGCCAATCTCGTCCTCTG
TTCCGTTGACGTCCGAGTTTCTCCAAGGAAACGAGGAAGAGGACGTGGCAAGTCTAATC
GATGGCTCAAGAAGTCTCAAAATGTTCCAATTCGATAGAAGTCTGTTGATACCAATTTGG
CAGGTGTAGGAACGTCTGATTTTGTGGTGCCAACTTACACCTCATGTACTGATCGTCA
ACGCCGAGAGGATGTGACGATGAAGATAATGACATTCTCTCAACAAGGATCTCGTGCTA
TCTGCATCCTTTAGCTAATGGTCCCATCTCCAATGTTACGCTTCGTCAATCTATGACAT
CCGGTGGTACTCTAACTTATGAGGGTCGTTTGTAGATTCTCTCTTTGACGGGTTCTGTTTA
TGCAAAATGACTCTGGAGGAACCTGAAGTAGAGCTGGTGGTATGAGTGTGCTTGCCTGCGAG
GACCAGATGGTCTGTCTTTGGTGGAGGACTCGTGGTCTCTTTCTTGCTGCTGGTCTCTG
TCCAGGTAATGGTAGGGACTTTTATAGCTGGTCAAGAGCAGTCAAGCTGGAGCTAGCAA
AAGAAAGACGGCTAAGATTTGGGGCTCAACCATCTTCTATCTCCTTTAACATATCCGCAG
AAGAACGGAAGGCGAGATTCGAGAGGCTTAACAAGTCTGTTGCTATTCTGCAACCAACCA
CTTCATACACGCATGTAAACACACAATAATGCGGTTACAGTTACTATACAACTCGGTTA
ACCATGTCAAGGATCCCTTCTCGTCTATCCAGTAGGAGGAGGAGGAGGTGAGAGGTAG
GAGAAGAAGAGGGTGAAGAAGATGATGATGAATTAGAAGGTGAAGACGAAGAATTCCGAG
GCGATAGCCAATCTGACAACGAGATTCCGAGCTGATGATGATCATAACGTTTCTTTTCGC
GGATTGTTAGGTTTGATGGATTTCAGATTTTGGTTGATTGTTTTATTAAACAGAAATG
TTTGAAGCTGCTATCTTTAGGTTCCCATCTCTTGTGATTGTTGAGTATCCTTGTGTA
AACAACTTACTGTTGCAAACTCTCTTCAAAAAGTTTCACTTTGCTTTCCCA

>G1399 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEEREGTNINNNITSSFLKQHEAAASDGGYSMDPPRPENPNPFLVPPTTVPAATVA
AAVTENAATPFSLTPTENTSABQLKKRGRPRKYNPDGLVVTLSMPPISSSVPLTSEF
PPRKRGRGRGKSNRWLKSQMFQFDRSPVDNLAVGTADFVGANFTPHVLIVNAGEDVT
MKIMTFSQQGSRAICILSANGPISNVTLRQSMSTSGGTLTYEGRFEILSLTGSFMQNDSSG
TRSRAGMSVCLAGPDGRVFGGLAGLFLAAGPVQVMVGTFIAGQEQSLELAKERRLR
GAQPSSISFNISAEERKARFERLNKSVAIAPPTTSYTHVNTTNAVHSYYTNSVNVHKDPF
SSIPVGGGGGGEVEEGEEDDELEGEDEFEFGGDSQSDNEIPS*

>G1415 (60..680)

CCTTATCACTACCAAAAGTCGTACATAATATCACTTTTCGAGTTATCAACATCCGTACA
TGTCATCCATAGAGCCAAAAGTAATGATGGTTGGTGCTAATAAGAAACAACGAACCGTCC
AAGCTAGTTTCGAGGAAAGTTGTATGAGAGGAAAAGGTGGACCCGATAACGCGTCTTGCA
CTTACAAAGGTGTTAGACAACGCACCTTGGGGCAAATGGGTGCGTGGATCCGCGAGCCTA
ACCGAGGAGCTCGTCTTTGGCTCGGTACCTTCGACACCTCCCGTGAAGCTGCCTTGGCTT
ATGACTCCGCACTCGTAAGCTCTATGGGCCTGAGGCTCATCTCAACCTCCCTGAGTCCT
TAAGAAGTTACCTAAAACGGCGTCTGCTCCGGCGTCCAGACTACACCAAGCAGCAACA
CCGGTGGAAAAGCAGCAGCGACTCTGAGTCGCCGTGTTATCCAACGAGATGTCATCAT
GTGGAAGAGTGACAGAGGAGATATCATGGGAGCATATAACGTGGATTGCGCGTAATGG
ATGATTCTTCAATATGGGAAGAAGCTACAATGTCTGTTAGGATTTCCATGGGTTTCATGAAG
GAGATAATGATATTTCTCGGTTTGATACTTGTATTTCCGGTGGCTATTCTAATTGGGATT
CCTTTTCACTCCCACTTTTGGAGTGCTACTAGACTCTCTTAATTGTTAAGTTATCATATA
CAAACATACATATATACAAATATAGTCACCGTGAACCTAGGATATATATGTAATAAACA
CCAGTTACATGTACTTATATATGTGCACATCTATATATGTGGTTTGTCTGTATAGTGTGA
AAGCAGATTCTTACCATATCA

>G1415 Amino Acid Sequence (domain in AA coordinates: TBD)
MSSIEPKVMVMGANKQRTVQASSRKGCMRGKGGPDNASCTYKGVQRQTWGWVVAEIREP
NRGARLWLGTFDTSREAAALAYDSAARKLYGPEAHLNLPESLRSPKTAASSPASQTPSSN
TGGKSSSDSESPCSSNEMSSCGRVTEEISWEHINVDLPVMDSSIWEEATMSLGFVWVHE

GDNDISRFDTCISGGYSNWDSFHSPL*

>G1417 (32..1501)

TCATCTCTATCTATCTCTCTTTGTCTGCAAAATGGAAGAACATATTCAAGATCGCCGTGA
AATTGCGTTCTTACACTCAGGAGAATTTCTCCACGGAGATTCTGACTCAAAGGATCATCA
ACCGAACGAGTCTCCGGTGGAAACGTCATCACGAGTCGTCTATCAAAGAAGTTGATTTCTT
CGCTGCTAAAAGTCAGCCGTTTGTATCTTGGTCATGTGAGAACAACGACGATCGTTGGATC
ATCTGGTTTTAATGATGGATTAGGTTTGGTAAATTCATGTCTATGGAACATCAAGCAATGA
TGGCGATGACAAAACCAAACTCAAATTAGTAGACTGAAGTTGGAGCTAGAGAGGCTTCA
CGAGGAGAATCACAACTGAAGCATTTATTAGATGAGGTCAGTGAGAGTTACAACGACCT
CCAAAGAAGAGTTTTGTAGCAAGACAAACACAAGTGGAAAGTCTTCATCATAAACAACA
TGAGGATGTACCTCAAGCTGGTTCTCACAAGCTCTAGAGAACAGAAGACCAAGGATAT
GAACCATGAAACTCCGGCCACCACCTTGAAACGACGGTCTCCAGACGACGTGGATGGTCG
TGATATGCACCGAGGATCACCAAAAACCTCTCGAATAGACCAAAAACAAGAGTACTAATCA
TGAAGAACAACAAAACCTCATGATCAATTACCTTATAGAAAAGCTAGGGTTTCCGTTAG
AGCTAGATCTGATGCCACTACGGTAAATGACGGATGTCAATGGAGAAAATACGGTCAGAA
AATGGCGAAAGGGAATCCATGTCTCGCGCTTATTATCGTTGCACCATGGCCGTTGGATG
TCCTGTCCGTAAACAGGTCCAACGATGCCGGAGGATACAACATATCTTGACAACAACGTA
CGAAGGAAACCATTAACCATCCTCTTCCCCCGTCAGCCACAGCCATGGCTGCAACCACTC
CGCCGAGCAGCCATGCTCTTATCAGGCTCCTCCTCCAGCAACCTCCACCAAACTCTC
TAGCCCCCTCCGCCACGTCATCATCATCTTCTACCATAACTTCCCATACACCTCCACAAT
CGCAACACTCTCTGCCTCAGCTCCTTCCCCACCATAACCTTAGACCTCACCAACCCACC
TCGACCGCTACAACCGCCACCGCAGTTTCTAAGCCAGTATGGTCCCGCCGCGTTTACC
AAACGCTAATCAAATTAGGTCTATGAATAATAAACCAGCAGTTATTAATACCTAATTT
GTTTGGCCCAACAGCCCCACACGTGAATGGTCGATTTCAGTTAGGGCTGCGATTGCGAT
GGATCGGAACCTTCACGGCGGCACCTTGGCGCCGCGATCTCAAACATTCGAGAGGAGTAA
TAACGACAACAATAATAACTGATATTAATGATAACAAGGTGATGCAAAAAGTGGAGG
GAGTAGTAACGGAGATTCGCCACAGCTTCTCAGTCTTGACCACTTTCTCTACAACTA
ATTTTACTACCATTATTATATGTTATCTTATTATATATTACACACATATTATACATTA
TGCGTATCTTAAGTTTTTTTTTGGGGGCCATTATATATGAATGATATGGAGATCACTGAG
AGAGAGAGAGAGCTATTATGGGTTTTTTTTT

>G1417 Amino Acid Sequence (domain in AA coordinates: 239-296)

MEEHIQDRREIAFLHSGEFLHGSDSKDHQPNESPVRRHHESSIKVDFFAAKSQPFDLG
HVRTTTIVGSSGFNDGLGLVNSCHGTSSNDGDDKTKTQISRLKLELERLHEENHKLKHL
DEVSESYNDLQRRVLLARQTQVEGLHHKHQEDVPQAGSSQALENRRPKDMNHETPATLTK
RRSPDDVDGRDMHRGSPKTPRIDQNKSTNHEEQQNPHDQLPYRKARVSVRARS DATTVND
GCQWRKYGQKMAKNPCPRAYRCTMAVGCPVRKQVQRCAEDTTILTTTYEGNNHPLPP
SATAMAATTSAAAAMLLSGSSSSNLHQLTSSPSATSSSSFYHNFYPTSTIATLSASAPFP
TITLDLTNPPRPLQPPPQFLSQYGPAFLPNANQIRSMNNNNQQLLIPNLFGPQAPPREM
VDSVRAAIAMDPNFTAALAAAIISNIIIGGNNDNNNTDINDNKVDAKSGSSNGDSPQLP
QSCTTFSTN*

>G1442 (1..1293)

ATGGGAACAAGAGCAGAACGCAAGGAAGATTTTGTGGTGGGTTTGGATTTGGTGTGTA
GAAATTCGCATAAAGACGTTATGGTGTACCTCATCATCACTATTATCCATCATATTCA
TCACCTTCCTCTTCTTCTTGTGTACTGTCTGTCTGGTGTAGCGATCCCATGTTCTCT
GTTTCTAGCAATCAGGCTTACACTTCTTCTCACAGTGGTATGTTTACACCCGCCGTTCT
GTTTCTGTCTGTGTGACTGTAGCAGATCCTTTTTTCTCCTTGAGCTCTTCAGGGGAAATG
AGAAGAAGTATGAACGAAGATGCTGGTGCAGCTTTCAGCGAAGCTCAATGGCATGAGCTT
GAGAGGCAGAGGAATATATACAAGTACATGATGGCTTCTGTTCTCTTCTCCAGAGCTT
CTCACACCTTTCCCAAGAACCACCAATCAAACACTAACCCGGATGTAAGTGTGGCAGTG
GCGACAGAGGCTCAFTCAGCTGGGGATTGCTTCAAGCGCAAGCAATAACACGGCTGAT
CTGGAGCCATGGAGGTGCAAGAGAACAGATGGGAAGAAATGGAGGTGCTCTAGAAACGTG
ATTCCTGATCAGAAATACTGTGAGAGACACACACAAGAGCCGTCCTCGTTCAAGAAAG
CATGTGGAATCATCTACCAATCATCTCACCACAATGACATTCTGACGGCTAAGAATGAT
ACTAGCCAGCTTGTGAGAACTTATCCTCAGTTTTTACCGACAACCTATAAGCCAGATCCCT
GTGCTTTCTACTCTTCCGTCTGCCTCCTCTCCATATGATCACCACAGAGGACTGAGGTGG
TTTACGAAAGAAGATGATGCCATTGGAACCTTAAACCCGGAGACTCAAGAAGCTGTCCAG
CTGAAAGTTGGATCAAGCAGAGAGCTCAAACGGGGATTTCGATTATGATCTGAATTCAGG

CAGAAAGAGCCAATAGTAGACCAGAGCTTTGGAGCATTGCAGGGTCTATTAAGTCTAAAC
CAGACACCACAACATAACCAAGAAACAAGACAGTTTGTGTAGAGGAAAGCAAGATGAA
GCGATGGGAAGCTCTCTGACACTCTCAATGGCTGGAGGAGGCATGGAGGAAACAGAGGGA
ACAAACCAGCATCAGTGGGTTAGCCATGAAGGTCCATCATGGCTCTATTCAACAACACCA
GGTGGACCATTGGCTGAAGCACTGTGTCTCGGTGTCTCCAACAACCCAAGTTCTAGTACT
ACTACTAGTAGCTGCAGCAGAAGCTCAAGCTAA

>G1442 Amino Acid Sequence (domain in AA coordinates: 172-223)

MGTRAERKEDFVGFGFVGVVENS HKDVMVLPHHHYPSYSSPSSSLCYCSAGVSDPMFS
VSSNQAYTSSHSGMFTPAGSGSAAVTVADPFFSLSSSGEMRRSMNEDAGAAFSEAQWHEL
ERQRNIYKMMASVVPPELLTPFPKNHQSNTPDVTVAATGGSLQLGIASSASNTAD
LEPWRCRRTDGKKWRCNRNIPDQKYCERHHTKSRPRSRKHVESHSQSSHHNDIRTAKN
TSQLVRYTPQFYGGPISQIPVLSLTPSASSPYDHRGLRWFTKEDDAIGTLNPETQEAVQ
LKVGSSRELKRGFDYDLNFRQKEPIVDQSFALQGLLSLNQTPQHNQETRQFVVEGKQDE
AMGSSSLTSLMAGGMEETEGTNQHQVWSHEGPSWLYSTTPGGPLAEALCLGVSNPNSSST
TTSSCSRSSS*

>G1454 (86..1180)

CTAGTAGTGATGATATGATCGCTTCTTCTCCTACAATCTCAGAAAACCTCCGATCACGGTT
TTAGATATCTTCTACAACGGATACAATGGAGAGCACCGATTCTTCCGGTGGTCCACCACC
GCCACAACCTAACCTTCTCCTCCAGGCTTCCGGTTTACCCTACCGACGAAGAGCTTGTGT
TCACTACCTCAAACGCAAAGCAGCCTCTGCTCCTTTACCTGTGCGCATCATCGCGAAGT
CGATCTCTATAAATTTGATCCATGGGAACCTCCCGCTAAAGCATCGTTTGGAGAAACAAGA
ATGGTACTTCTTTAGTCCACGAGATCGGAAGTATCCAAACGGAGCAAGACCAACAGAGC
GGCGACTTCAGGTTTATGGAAAGCGACCGGTACAGATAAACCGGTACTTGCTTCCGACGG
TAACCAAAAGGTGGGCGTGAAGAAGGCACTAGTCTTCTACAGTGGTAAACCACCAAAAGG
CGTTAAAGTGATTGGATCATGCATGAGTATCGTCTCATCGAAAACAAACCAACAATCG
ACCTCCTGGCTGTGATTTCCGGCAACAAAAAACTCACTCAGACTTGATGATTGGGTGTT
ATGTAGAATCTACAAGAAGAACAACGCAAGTCGACATGTTGATAACGATAAGGATCATGA
TATGATCGATTACATTTTCAGGAAGATTCTCCTCGTCTTTATCAATGGCGGCTGCTTCTAC
AGGACTTCACCAACATCATATAATGTCTCAAGATCAATGAATTTCTTCCCTGGCAAATT
CTCCGGTGGTGGTTACGGGATTTTCTCTGACGGTGTAAACAGAGTATATACGACGGCGG
TGGCATGATCAACAATATTGGTACTGACTCAGTAGATCACGACAATAACGCTGACGTCGT
TGGTTTAAATCATGCTTCGTCGTCAGGTCCTATGATGATGGCGAATTTGAAACGAACTCT
CCCGGTGCCGTATTGGCCTGTAGCAGATGAGGAGCAAGATGCATCTCCGAGCAAACGGTT
TCACGGTGTAGGAGGAGGAGGAGGAGATTGTTTCAACATGTCTTCTCCATGATGGAAGA
GACTCCACCATTGATGCAACAACAAGGTGGTGTGTTAGGAGATGGATTATTCAGAACGAC
ATCGTACCAATTACCCGGTTTAAATTTGGTACTCTTCTTAATCAATGTGTTTCGCGCGG
GTGTGAAGAATTTTCCGGTGACAGTGAAGATTTTCTCCGATTGGTGGGGTCATTGTCAT
GCATTATATAATTTGAGATTTGTGTATATGTTTTGGGTTAATTAATTGGTCACAGGGGC
>G1454 Amino Acid Sequence (conserved domain in AA coordinates: 9-178)

MESTDSSSGPPPPQPNLPPGFRFHPTDEELVVHYLKRKAASAPLPVAIIAEVDLYKFDPW
ELPAKASFGEQEWYFFSPDRKYPNGARPNRAATSGYWKATGTDKPVLASDGNQKVGKK
ALVFYSGKPPKGVKSDWIMHEYRLIENKPNRPPGCDGFKNKNLRLDDWVLCRIYKNN
ASRHVDNDKDHMDIFYFRKIPLPSMAAATSTGLHQHHNVSRSMNFFPGKFSGGGYGIF
SDGGNTSIYDGGGMINNIGTDSVDHDNADVVGLNHASSSGPMMANLKRTPVPYWPVA
DEEQDASPSKRFHVGVGGGGDCSNMSSSMMEETPPLMQQGGVLGDGLFRTTSYQLPGLN
WYSS*

>G1459 (1..1272)

ATGATGAAAGGTCTGATTGGGTATAGATTTAGTCCGACGGGAGAGGAAGTGATCAACCAT
TACCTAAAGAACAACCTTCTGGGTAAGTATTGGCTCGTTGATGAAGCTATTAGCGAGATC
AACATCTTGAGTCACAAACCCAGCAAGGATTTGCCTAAGTTAGCTAGGATCCAATCGGAA
GATCTTGAATGGTATTCTTCTCTCCGATTGAGTACACGAACCCGAATAAGATGAAAATG
AAGAGGACGACAGGTTCTGGGTTTTGAAACCTACTGGTGTGATCGGGAAATTAGGGAT
AAAAGAGGAAATGGTGTGTGATAGGGATTAAGAAGACGCTTGTGTACCATGAAGGTAAG
AGTCCTCATGGAGTTAGAACTCCTTGGGTTATGCACGAGTATCACATCACTTGCTTGCT
CATCATAAGAGGAAATATGTTGTCTGCCAAGTAAAGTATAAGGGTGAAGCTGCAGAAATT
TCATATGAGCCCAAGTCCCTCTTTGGTATCCGATTTCGCATACCGTCATAGCGATTACCGGA
GAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAGGTAAAGAAAATCTCTTGGGTATGTCT

GTAGATGATTTGATAGAACCAATGAACCAACAAGAGGAGCCACAAGGTCCTCACTTAGCT
CCGAATGATGATGAGTTTATACGTGGATTGAGGCATGTTGATCGAGGGACGGTTGAATAT
TTGTTTGCCAATGAAGAAAACATGGATGGTTTGTCTATGAATGACTTGAGAAATCCCAATG
ATCGTCCAACAAGAGGATCTCTCTGAGTGGGAGGGATTTAACGCAGACACCTTTTTCAGC
GACAACAACAATAACTATAACCTTAACGTGCATCATCACTAACGCCTTACGGCGATGGC
TATTTGAATGCATTTTCGGGTTATAACGAAGGGAATCCTCCGATCACGAATTAGTGATG
CAAGAGAACCGCAACGATCACATGCCAAGGAAACCTGTGACAGGGACCATTGATTATAGC
AGCGATAGTGGCAGTGATGCTGGATCCATATCTACAACGGTGAAACAAGAAATCCCAAGA
GCTGTTGATGCACCCATGAACAATGAGTCATCTTTGGTGAAAACAGAGAAGAAAGGCTTG
TTTATTGTAGAGGACGCAATGGAGAGAAACCGCAAGAAACACGATTTATCTATCTCATG
AAGATGATCATAGGCAACATCATATCGGTTTTACTACCCGTCAAAAGATTGATCCCGGTG
AAGAAGTTATGA

>G1459 Amino Acid Sequence (conserved domain in AA coordinates:10-152)

MMKGLIGYRFSPTGEEVINHYLKNKLLGKYWLVEAISEINILSHKPSKDLPLKARIQSE
DLEWYFFSPIEYTNPNKMKMKRTTSGSFWKPTGVDRIRDKRNGVVGIGIKKTLVYHEGK
SPHGVRTPWVMHEYHITCLPHHKRKYVVCQVKYGEAAEISYEPSPLVSDSHTVIAITG
EPEPELQVEQPGKENLLGMSVDDLI EPMNQEEPPQGPLAPNDDEFIRGLRHVDRGTVEY
LFANEENMDGLSMNDLRIPMIVQQEDLSEWEGFNADTFFSDNNNNYNLNVHQLTQPYGDG
YLNAFSGYNEGNPPDHELVMQENRNDHMPRKPVGTIDYSSDSGSDAGSISTTVKQEI PR
AVDAPMNNESLSLVKTEKKGLFIVEDAMERNRKKPRFIYLMKMIIGNIISVLLPVKRLIPV
KKL*

>G1460 (87..995)

CGTCGACCTTCACTCAAAACCTAATCCCGGGAACCCGGGAATTTTGATCATTGTTTCT
TTTCGATCTGTTTCTATTTTAAAAAGATGATGAAAGATCCGACTGGGTATAGATTAGTC
CGACGGGAGAGGAAGTGATAAACCATTAACCTAAAGAACAAAATTTGCGGTAAGACTTGGC
TCGTTGATGAAGCCATTAGCGAGATCAACATCTTGAATCACAACCCAGCAAGGATTGTC
CTAAGTTAGCTAGGATCCAATCGGAAGATCTTGAGTGGTACTTTTCTCTCCGATTGAGT
ACACGAACCCGAATAAGATGAAAATGAAGAGGACGACAGGTTCTGGGTTTTGGAAACCTA
GTGGTGTGATCGGAAAATTAGGGATAAAAAGAGGAAATGGTGTGTGATAGGGATTAGA
AGACGCTTGTGTACCATGAAGGTAAGAGTCCTCATGGAGTTAGAACTCCTTGGGTTATGC
ACGAGTATCACATCACTTGCTTGCCCTCATCATAAGAGGAAATATGTTGTCTGCCAAGTAA
AGTATAAGGGTGAAGCTGCAGAAATTTTCATATGAGCCAAGTCCCTCTTGGTATCCGATT
CGCATACCGTCATAGCGATTAAACGGAGAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAG
GTAAAGAAAATCTCTTGGGTATGTCTGTAGATGATTTGATAGAACCAATGAACCAACAAG
AGGAGCCACAAGGTCCTCACTTAGCTCCGAATGATGATGAGTTTATACGTGGATTGAGAC
ATGTTGATCGAGAGCCGGTTGAATATTTGTTTGCCAATGAAGAAAACATGGATGGTTTGT
CTATTATGAATGACTTGACAATCCCAATGATCGCCCAACAAGAGGATCTCATTCTCTCTG
AGTGGGAGGGATTATCGCAGCCACCTTTTTCAGCGACAACAACAATAACAATAACCTTA
ACGTGCATCAACTAACGTTCTTCTTACCAGGATGATTATCAGAATGCATTTTGGGTTACA
ACGGAGCGNCCGCT

>G1460 Amino Acid Sequence (domain in AA coordinates: TBD)

MMKDPTGYRFSPTGEEVINHYLKNKILGKTWLVDEAISEINILNHKPSKDLPLKARIQSE
DLEWYFFSPIEYTNPNKMKMKRTTSGSFWKPSGVDRKIRDKRNGVVGIGIKKTLVYHEGK
SPHGVRTPWVMHEYHITCLPHHKRKYVVCQVKYGEAAEISYEPSPLVSDSHTVIAING
EPEPELQVEQPGKENLLGMSVDDLI EPMNQEEPPQGPLAPNDDEFIRGLRHVDREPVEY
LFANEENMDGLSIMNDLTIPMIAQQEDLILSEWEGFIAATFFSDNNNNNNLNVHQLTSL
PG*

>G147 (37..672)

AAATCATCAGATAGAAGGAAATATTCTGATTGAGAGATGGCTCGTGGAAGATTGAGCTT
AAGAGGATTGAGAACECGGTTACAGACAAGTGACTTTTTCAGAGAGGAGAACTGGTCTT
CTCAAGAAGGCTAAGGAGCTCTCTGTGCTCTGTGATGCCGAGATCGGTGTTGTGATCTTC
TCTCCTCAGGGCAAGCTCTTTGAGCTCGCTACTAAAGGAACAATGGAGGGAATGATTGAT
AAGTACATGAAGTGTACTGGTGGTCTGTTCTTCTCTGCTACTTTTACTGCTCAA
GAACAACCTTCAACCACCAATCTTGATCCGAAAGATGAGATCAACGTGCTTAAGCAAGAG
ATTGAGATGCTTCAGAAAGGATAAGCTATATGTTTGGAGGAGGAGATGGGGCTATGAAT
CTTGAAGAACTTCTTTGCTTGAGAAGCATCTTGAGTATTGGATTTCTCAGATTTCGCTCT
GCTAAGATGGATGTTATGCTTCAAGAAATTCAGTCATTGAGGAACAAGGAAGGAGTCCTC

AAAAACACCAACAAGTATCTCCTCGACAAGATAGAGGAAAAACAACATAGCATATTAGAT
GCTAACTTCGCAGTCATGGAGACAAACTATTCCTATCCGCTAACAAATGCCAAGTGAAATA
TTTCAGTTCTAGACCATAGGGTATTTGAAGACTATGTCTCACGAATTTAAATAACCTTGG
TAAGTATAATATAGTGTGTAAATCACACATAATTAATAAAGCCTGTGGAACCTCGC
TAGGCAGTTGAAAATCTATCCGTATGTTTTATCCTCTTGTTTTACATTTGTTGGTGTGAA
GATGAAATGACTGCAAGTGTGGTGTGTACTTATAACTCTTCTACTTTCTATCTATGTTT
TGAATTTATGGATT

>G147 Amino Acid Sequence (domain in AA coordinates: 2-57)
MARGKIQLKRIENPVHRQVTFCKRRTGLLKKAKELSVLCDAEIGVVIFSPQGLFELATK
GTMEGMIDKYMKCTGGGRGSSATFTAQEQLOPPNLDPKDEINVVKQEIEMLQKGISYMF
GGGDGAMNLEELLLEKHLEYWISQIRSAKMDVMLQEIQSLRNKEGVLKNTN KYLLDKIE
ENNNSILDANFAVMEITNYSYPLTMPSEIFQF*

>G1471 (1..735)

ATGGAGAACCAATCTATGTCTTCATCAAGCTCCTCCACACACAAACATGATCAAAAACCTC
AAAAGTTCCGTTGTGGCCATGGAGGTCTTGGAGGAGAAGGAGACAGTGAACAATCCGCCC
CAGTATTATAAAGATCTACATCTGTTACTTGTGCAAGAGAGCGTTCCCAACCCCTCAT
GCCCTTGGCGGTACGGAACCCACCAAGGAGGACCGAGAATTGGAGAGGCAACAGATC
GAGTCAAGGCTTTCTAACAAAGACAAGTCTAACTTGCTCTTGGTGGGTCTTCACAAGAT
GTTTTATCAAATGATAATCACCTTGGACTCTCTCTTGGTCCATTGAAGTCCATAGAAGGT
AGCAGCAGCAGCAACAACGTTAACCATTGCTTAATGTTGGAGTCCCTAGAGGAACCACA
GATATGAACATGAACAACATATAGCTCACATGCTTTATCAACTGATGATATTAATCTTGAT
CTTACTCTTGGTCCATCTAAGTCCATAGGAGATAGCAACAATATCATTAATAACAACACT
AATCATCTCTCGATGGGAATCTGATCATTCCTCGTTCCTCGTGTGTCTAGATACCAT
TTTGTGTGCTGGGAACCCCTTGATTCAATCTCTAGAAACATTCTCCTCTCTATTACTTTT
CCTCATCTAAACATCAATCTTCTCATGATTCTGTTTTCTTTACAAGAGAATGGTTCGGGC
TCTAGTCACTCATAA

>G1471 Amino Acid Sequence (domain in AA coordinates: 49-70)
MENQSMSSSSSTHKHDQKLKSSVVAMEVLEEKETVNNPPQYYNKIYICYLCKRAFPTPH
ALGGHGTTHKEDRELERQIESRLSNKDKSNLLFGGSSQDVLSDNHLGLSLGPKLSIEG
SSSSNNVNP LLNVGVPRGTTDMNMNYSHALSTDDINLDTLGPSSKISGDSNNIINNNT
NSSFDGNLIIPVRPRVSRHYFVAGNPLDSISRNI PPSTP PHLNINLSHDSFSLQENGSG
SSHS*

>G1475 (1..645)

ATGAAGAGAACACATTTGGCAAGTTTTAGTAACAGAGACAAAACCCAAGAAGAAGAAGGA
GAAGACGGTAATGGTGACAACAGAGTCATCATGAATCACTACAAGAATTACGAAGCTGGG
CTGATCCCATGGCCTCCCAAGAATTACACTTGCAGCTTCTGCAGGAGAGAGTTTCAATCT
GCTCAAGCACTTGGAGGCCACATGAATGTTCATAGAAGAGACAGAGCAAAACTCAGGCAG
ATCCCTTCTTGGCTCTTTCGAACCTCACCACCACACACCTATTGCAAACCCTAACCCCTAAT
TTTAGCTCTTCTTCTTCTTCTTCAACAACAACAGCTCATCTTGAGCCTTCCCTAACCAAC
CAGAGATCCAAAACCACTCTTTCTTCTTCTGCCCCGTTTGATCTTTTGGACAGTACTACT
AGCTATGGAGGTTTGATGATGGACAGAGAGAAGAACAAGAGCAATGTATGTAGCAGAGAG
ATCAAGAAAAGTGCCATCGATGCATGTCATTAGTAAGATGTGAGATAAGCCGTGGGGAT
CTGATGAATAAGAAAGATGATCAAGTCATGGGTTGGAGCTTGGGATGAGTTTGAGGAAT
CCCAACCAAGTTCTTGATTTGGAGCTTCTGACTAGGCTACCTCTAA

>G1475 Amino Acid Sequence (domain in AA coordinates: 51-73)
MKRTHLASFSNRDKTQEEEGEDGNGDNRVIMNHYKNYEAGLIPWPPKNYTCFRCRREFRS
AQALGGHMNVHRRDRAKLRQIPSWLFEPHHHTPIANPNPNFSSSSSTTTAHLEPSLTN
QRSKTTFPFSARFDLDTSTSYGGLMMDREKNKSNVCSREIKKSAIDACHSVRCEISRGD
LMNKKDDQVMGLELGLMLRNPNQVLDLELRGLYL*

>G1477 (1..606)

ATGTTGTCTCGGACTCGAATTACGCTAGTGATATTAGCGACGATGCCTCCGCCACCGBA
TCGATAGAGAATCCTATATACAAATGCAAGTATTGTCTAGGAAGTTGATAAAACACAA
GCATTAGGTGGTCATCAAAATGCACACAGAAAGGAGAGAGAGGTGAAAAACAACAAAA
GCATTTTTGGCGCATTTGAACGACCAGAACAGATCTTTACGCGTACTCGTATTCTGTAT
CATCATTCATTTCTTAACCAATACGCACCTCCACCGGGATTGTAACAGCTCAGTACAAA
GTTGATAGATCATACAAGATGTCATGGTCTACAACCAATATGTGGGATCCTCAAGCTCT
AGCTTTGCAGGACTACAAAGTGACCCAAGTCAAGGAATGAACCAGGATTGGACCTTTACC

GGGATCCCATTCTACCCCAATCTCAACCTCAACCACTATCGTCACCAATATGTTTGGAT
CTTTGCCTTGGCATTTGGTAGCTCCCAAACCCAACCACACCTCAAGAACCAATGATGCA
ACAGAAGAGATGGATGCTGAGAAAGAAAATGATGGTTCTTCCCTTTCTCTCACTCAA
CTGTGA

>G1477 Amino Acid Sequence (domain in AA coordinates: 29-48)
MLSSDSNYASDISDDASATGSIENPIYKCKYCPRKFDKQALGGHQNHRKEREVEKQKQ
AFLAHLNRPEPDLAYSYSYHHSFPNQYALPPGFEPQYKVDRSYKMSMVYNQYVVGSSSS
SFAGLQSDPSQGMNQDWTFTGIPFLPQSQPQLSSPICLDLCLGIGSSQTQPQPQEPNDA
TEEMDAEKENDGSSLSLSLKL*

>G1487 (1..1020)
ATGGAACAAGCCGCGTTGAAGAGCAGCGTCAGGAAAGAGATGGCTCTCAAAACGACTTCT
CCGGTTTACGAAGAGTTTCTTGCCGTCAACACCGCTCAAAATGGCTTTTCCGTCGACGAT
TTCTCTGTAGACGACTTGCTTGACTTGTCAAACGATGACGTTTTTGGCCGACGAAGAACT
GACCTCAAGGCTCAACATGAGATGGTCCGTGTTTCTCTGAGGAACCCACGACGACGGA
GACGCTCTTCGCCGAGCAGCGATTCTCCGGCTGTGACGACTTTGGTTCTCTCCCTACA
AGCGAATCTCTCTTCGGCGGATGATTTAGCGAACCTTGAGTGGCTCTCTCATTTCTGTG
GAGGACTCTCTTCACGGAATATTCCGGTCCAAACCTCACCGAACCCCGACTGAGAAACCG
GCGTGGTTAACGGGTGACCGGAAACATCCTGTGACTGCAGTCACGGAAGAGACCTGTTTC
AAATCCCCTGTTCCGGCTAAAGCCCGTAGCAAACGTAACCGCAATGGCCTCAAGGTCTGG
TCGCTTGGTTCTGCTCCTCTCCGGTCTCTCGTCCGGTTCGACCTCTCTCCTCTCT
TCGGGTCTCTCCAGCCCGTGGTTCTCCGGCGCTGAGCTGCTCGAGCCTGTGGTCACGTCA
GAGAGGCCACCGTTTCCCAAGAAGCATAAGAAAAGGTGAGCCGAGTCTGTTTTCTCCGGT
GAGCTGCAGCAGCTGCAACCTCAGCGAAAGTGCAGCCACTGCGGCGTTTCAAGAACTCCG
CAGTGGAGGCCCGGCAATGGGAGCCAAAGACCTGTGCAATGCGTGCCTGTCTCCGGTAC
AAGTCCGGTAGGTTGCTACCGGAATACAGACCCGCTTGTAGCCCGACATTCTCGAGTGAG
CTGCACTCGAACCAACACCGGAAAGTCATAGAGATGAGGCGGAAGAAGGAGCCAACAGT
GACAACGAAACCGGTTTAAACAGCTGGTTTCACTCCCAAGCTGTACCAAGTTTTTGA
>G1487 Amino Acid Sequence (domain in AA coordinates: 251-276).
MEQAALKSSVRKEMALKTTSPVYEEFLAVTTAQNGFSVDDFSVDDLLDLSNDDVFADDEET
DLKAQHEMVRVSEEPNDDGDLRRSSDFSGCDDFGLPTSELPLPADDLANLEWLSHFV
EDSFTEYSGPNLTGTPTKPAWLTDGRKHPVTAVTEETCFKSPVPAKARSKNRNGLKVV
SLGSSSSSGPSSSGSTSSSSSGPSSPWFSGAELLEPVVTSEPPFPKHKHRSAESVFSG
ELQQLQPPQRKCSHCGVQKTPQWRAGPMGAKTLCNACGVRYKSGRLLPEYRPACSPFSSSE
LHSNHHRKVIEMRRKKEPTSDNETGLNQLVQSPQAVPSF*

>G1492 (149..919)
AATCCCAACCCACACACCTCTCAAATCCTCCTCTCCTCGTTTCTCTTTCTCTCCTCTTCA
CAGAACCAAAACATATCAAACCTTTTTTCTCTTGGGTTTAAAGTAAAAATCGAATCTTTG
TGTGCGTTTTTGGGTTCTTGAACGATATGGGTAAGTCTAGTGGTAGAAATGGTAACGG
AAGCTTTAACGGCAATAAATTTACCGGAGTTAGACCTTACGTACGGTCTCCAGTTCCACG
GCTTAGATGGACGCCGATCTTACCGTTGTTTCGTTACGCCGTCGAGATTCTCGGTGG
TCAACACCGAGCAACACCAAACTTGTTCTTAAGATGATGGATGTGAAGGGACTTACCAT
TTCACATGTCAAAAGCCACCTTCAAGATGTATAGAGGAGGTTCAAAGCTCACTTTGGAGAA
ACCAGAAGAAAGCTCATCATCTTCAATAAGAAGAAGACAGACAGTGAAGAAGATTATTA
TCTTCATGACAACTTGTCTTTACACACAAGGAATGATTGTCTTTTGGGTTTCACTCTTT
TCTCTTTTCTTACATTCTTCAATTAGAGGAGGAGGAGGAGGAAGAACAAGAGCAGCA
GACTTCAGAGTCTGGTGGTTATGATGATGATGCTGACTTCTTACATCAAGAAGATGAA
CGATACGACGACGTTTTTGTACATCATTTCCCAAGGGAACAGAGGAGTGGCGGGAACA
AGAACACGAAGAAGAAGAAGATTGTCTGTTGTCTCTGCTGTTAAATCATCATATTG
GAGAAGCAATGATCATCGGTGGTGAGCGAAACGAGTGAAGCAGCAGTCTCGACTTGTTC
AGCACCATTCTGATCCAAAGATTGCTTTGGTTCTTCAAAGATTGATCTTAATCTGTCAAT
TTCTCTCCTCGGTAGCTAAATAAGTTATGCAAGATTAGGTTTCAAGAACTATTCCGAT
GTGTTTTTGAAGTATGATATTGAATGTTAGTAGAGAACTAGAAATGAAGTTTAGAT
AAATTATCAACGCAGCGTTTTGATCGCCTTTGAACGGAAATTAACAAA

>G1492 Amino Acid Sequence (domain in AA coordinates: 34-83)
MGKSSSRNGNSFNGNKFHGVPRPVRSVPRLRWPDLHRCFVHAVEILGGQHRATPKLV
LKMMDVKGLTISHVKSHLQMYRGSKLTLEKPESSSSSIRRRQDSEEDYVYLDNLSLHT
RNDCLLGFHSFPLSSHSSFRGGGGGRTKEQQTSESAGYDDADFLHIKMDNTTFLSHH

FPKGTEEWREQEHEEEEEEDLSLSLSLNHHHWSRNGSSVVSETSEAAVSTCSAPFVSKDCF
GSSKIDLNLISISLLGS*

>G1531 (1..666)

ATGTGTGAGTCAAGCAACAAAGTCAGAGTATCGCCATACCCGCTTCGGTCTTCGAGGACC
GACAAACACAAGGCGTCAGAGTCGCCTATTGAGACAGGTTGGGAGGATGTGCGTGGATGT
CATCCTTACATGTGCGATACGAGTGTTCTGTCACCTCCAATTGTTTCAAGCAGTTCGCGAGA
AAAACCATAAAAAGCGCCTATACCCCAAGACCTTACATTGTCCTCTCTGTAGAGGTGAA
GTATCCGAGACGACAAAGGTGACGAGCACTGCAAGAAGATTATGAATGCTAAACCGAGG
TCTTGCTCCGTAGAGGATTGCAAATTCTCTGGGACGTTTTCTCAGCTTACTAAGCACTTG
AAAACGAGCATCGCGGTATTGTGCCACCAAGGTCGATCCACTGAGACAACAGAGATGG
GAAATGATGGAGAGACATTCTGAATACGTTGAACTCATGACTGCAGCTGGGATTTCGCGT
ATGGCTGAGGTGATGCAACAACAGCTTCCCCAGGATCAGAATCATCTCATGTGTTCAA
GTGACCGTTAATGGAACCATATGGAATCTAATTGATCCGAGTCAGGGAAGGAATGGATTA
GGCATACCAACTATAGCGCAATGCAGTTTGTACCATTAAAGCATAAATCACAGTAGAACT
CTGTGA

>G1531 Amino Acid Sequence (domain in AA coordinates: 41-77)

MCESSNKVRVSPYPLRSSRTDKHKASESPIETGWEDVRGCHPYMCDTSVRHSNCFKQFRR
KTIKKRLYPKTLHCLCRGEVSETTKVTSTARRFMNAKPRSCSVEDCKFSGTFSQLTKHL
KTEHRGIVPPKVDPLRQQRWEMMERHSEYVELMTAAGISRMAEVMMQQQLPQDQNHPIVFO
VTVNGTIWNLDIPSQGRNGLGITNYSAMQFVPLSINHRTL*

>G1540 (122..997)

atctctttactaccagcaagttgtttcttgtaacttcaaacttctctttctctgttc
ctctctaagtcttgatcttatttaccgttaactttgtgaacaaaagtcgaatcaaacaca
catggagccgccacagcatcagcatcatcatcaagccgaccaagaaagcggcaacaa
caacaacaagtcgggctctgggtgttacacgtgtcgccagaccagcacgaggtggacacc
gacgacggagcaaatcaaaatcctcaaagaactttactacaacaatgcaatccggtcacc
aacagccgatcagatccagaagatcactgcaaggctgagacagttcggaaagattgaggg
caagaacgtcttttactgggtccagaaccataaggctcgtgagcgtcagaagaagagatt
caacgggaacaaacatgaccacaccatcttcacccaactcgggttatgatggcggctaa
cgatcattatcatcctctacttaccatcatcacggtgttcccatgcagagacctgctaa
ttccgtcaacgttaaaacttaaccaagaccatcatctctatcatcataacaagccatatcc
cagcttcaataacgggaatttaaatcatgcaagctcaggtactgaatgtggtgtgttaa
tgcttctaattggctacatgagtagccatgtctatggatctatggaacaagactgttctat
gaattacaacaacgttaggtggaggtgggcaaacatggatcatcattactcatctgcacc
ttacaacttcttcgatagagcaaagcctctgtttggtctagaaggtcatcaagacgaaga
agaatgtggtggcgatgcttatctggaacatcgacgtacgcttccctctctccctatgca
cgggtgaagatcacatcaacggtggttagtggtgccatctggaagtatggccaatcggaagt
tcgcccttgcgcttctcttgagctacgtctgaactagctcttacgcgggtgtcgctcggg
attaaagctcttccctctctctctctctctctctgactcgtatgttcacaactatgcttcgc
tagtgattaatgatgcagttgttatattagtagttaactagttatctctcgttatgtgta
atttgtaattactagctaagtatcgtctaggtttaattgtaattgacaaccggttatctc
tatgatgaataagttaaatttatatat

>G1540 Amino Acid Sequence (domain in AA coordinates: 35-98)

MEPPQHQQHHHQADQESGNNNNKSGSGGYTCRQTSTRWTPTEQIKILKELYNNNAIRSP
TADQIQKIKITARLRQFGKIEGKNVYWFQNHKARERQKKRFNGTNMTTPSSSPNSVMMAAN
DHYHPLLHHHHGVPMQRPANSVNVKLNQDHHLYHHNKPYPSPFNNGNLNHASSGTECGVVN
ASNGYMSSHVYGSMEQDCSMNYNVVGWWANMDHHYSSAPYNFFDRAKPLFGLGHHQDEE
ECGGDAYLEHRRTLPLFPMHGEDHINGGSGAIWKYQSEVRPCASLELRN*

>G1544 (1..2178)

ATGTCTCAGTCAAACATGGTACCAGTGGCTAACAAACGGAGACAACAACAGACAACGAA
AACAAACAACAACAACAACAATGGTGGAACTGACAACACTAATGCTGGAAATGATTCT
GGAGATCAAGATTTTCGACAGTGGGAATACCTCAAGTGGCAATCATGGAGAAGGGTTGGGA
AACAAATCAAGCTCCTCGTCATAAGAAGAAAAATACAATCGTCACACCCAACCTCAGATT
TCGGAGATGGAAGCTTTCTTCAGAGAGTGTCTCACCCAGATGACAAACAAAGGTACGAC
CTTAGCGCTCAATTGGGATTGGACCCCTGTTAGATCAAATTCGTTCCAGAACAAACGC
ACTCAAAACAAGAATCAACAAGAACGCTTTGAGAACTCAGAACTTCGGAATCTGAACAAC
CACCTTAGGTCTGAAAATCAGCGGTTACGAGAAGCTATTCATCAAGCCTTATGCCCTAAG

TGTGGAGGCCAACTGCAATTGGCGAAATGACCTTCGAAGAGCACCATCTTCGCATCCTC
AACGCTCGTTTACTGAAGAGATCAAGCAACTTTCGTGACAGCGGAAAAGATATCAAGG
CTTACGGGGATACCAGTAAGGAGCCATCCCCGTGTGTCTCCTCTAATCCTCCTCCAAAT
TTCGAGTTCGGGATGGGATCTAAGGGAAATGTCGGAACCACTCGAGGGAAACCACTGGA
CCTGCAGATGCTAATACCAAGCCGATCATCATGGAGTTGGCATTGGAGCCATGGAGGAG
CTCTTGGTGTATGGCTCAAGTGGCTGAACCACTGTGGATGGGAGGATTTAATGGCACTAGC
TTAGCTTTGAACCTTGGATGAATACGAAAAGACGTTTCGCACGGGTCTCGGTCCCTAGACTT
GGCGGGTTTCGAACCGAGGCATCCAGGGAACTGCACTCGTGGAATGTGTCTCTACTGGC
ATTGTTGAAATGCTCATGCAAGAGAATCTGTGGTCAACAATGTTGCCGGAATTGTGGT
AGAGCCAGGACTCATGAACAGATAATGGCTGATGCTGCTGGAACCTTCAATGGAAATCTC
CAAATAATGAGTGCTGAGTACCAAGTGCTTTCCCCGCTAGTCACAACCCGCGAAAGCTAC
TTCGTCCGCTACTGTAAGCAACAAGGAGAGGGTTGTGGGCGGTGGTGCATATTTCCATC
GACCATCTCCTCCCAAACATCAACCTAAATGTGCGCCGCGACCTCTGGATGTCTGATT
CAAGAAATGCATAGTGGTTACTCCAGGTTACATGGGTGGAACATGTGGAAGTAGATGAT
GCAGGAAGTTACAGCATCTTTGAGAAATTAATCTGTACTGGTCAAGCTTTTGTCTGCTAAC
CGCTGGGTTGGTACATTTGGTACGCCAGTGTGAGCGGATATCTAGCATCTTGTGCACAGAT
TTTCAATCTGTGCTGATTCGGGTGATCACATAACGCTAACTAACCATGGAAAGATGAGCATG
CTGAAGATAGCTGAGCGGATTGCGAGAACCTTCTTTGCTGGAATGACCAATGCGACGGGG
TCTACAATATTTTCTGGTGTGTAAGGAGAAGATATCAGAGTGATGACAATGAAGAGCGTG
AATGATCCAGGAAAGCCTCCCGGTGTATTATTTGTGCAGCCACTTCTTTTGGCTTCTCT
GCTCCTCCTAACACTGTCTTTGACTTCTCTCAGAGAGGCTACTCACCACACAATTTGGGAT
GTTCTCTGCAACGGAGAGATGATGCACAAGATAGCAGAGATTACGAATGGGATAGACAAA
AGGAAGTGTGCAAGTTTACTCCGGCATGGACACACTAGCAAGAGCAAGATGATGATAGTT
CAAGAGACTTCTACTGACCCCAACAGCTTCATTTGTGCTTTATGCGCCTGTTGATATGACA
TCAATGATATTACTCTCCATGGAGGTGGTGTATCCTGACTTTGTGGTGATCCTGCCTTCT
GGTTTTGCTATTTTCCAGATGGTACGGGTAAAGCCTGGAGGAAAAGAGGAGGATCACTT
TTGACCATTTCTTCCAAATGCTGGTTGAGTCAGGTCCTGAGGCTAGGCTGAGTGTTAGC
TCTGTTGCAACTACTGAGAATCTGATTCTGTAACCGTGCGGAGGATCAAAGATTTGTTT
CCTTGTGCACTGCTTGA

>G1544 Amino Acid Sequence (domain in AA coordinates: 64-124)

MSQSNMVPVANNGDNNNDNENNNNNNGGTDNTNAGNDSGDQDFDSGNTSSGNHGEGLG
NNQAPRHKKKKYNRHTQLQISEMEAFFRECPHPDDKQRYDLAQLGLDPVQIKFWFQNK
TQNKQQRERFENSELRNLNHLRSENQRLREAIHQALCPKGGQTAIGEMTFEEHHLRLIL
NARLTEEIKQLSVTAEKISRLTGIPVRSHPVSPNPNNPFEGMGSKGNVGNHSRETTG
PADANTKPIIMELAFGAMBEELLVMAQVAEPLWMGGFNGTSLALNLDEYEKTFRTGLGPRL
GGFRTEASRETALVAMCPTGIVEMLMQENLWSTMFAGIVGRARTHEQIMADAAGNFNGNL
QIMSAEYQVLSPLVTTRESYFVRYCKQQGEGWLAVVDISIDHLLPNINLKCRRRPSGCLI
QEMSGYSKVTWVTEHVEVDDAGSYSIFEKLICTGQAFANRWGTLVRQCERISSILSTD
FQSVDSGDHITLTNHGKMSMLKIAERIAITFFAGMTNATGSTIFSGVEGEDIRVMTMKS
NDPGKPPGVIIICAATSFWLPAAPPNTVFDLREATHRNWDVLCNGEMMHKIAEITNGIDK
RNCASLLRHGHTSKSKMMIVQETSTDPTASFVLYAPVDMTSMIDITLHGGGDPDFVILPS
GFAIFPDGTGKPGGKEGGSLLTISFQMLVESGPEARLSVSSVATTENLIRTTVRRIKDLF
PCQTA*

>G156 (39..755)

AGGAAGAGGGAGCCACTCATAAGAGGAAGAAGAGAGAGATGGGTAGAGGGAAAGATAGAGA
TAAAGAAGATAGAGAATCAGACGGCGAGGCAAGTGACCTTCTCCAAGAGAAGAACTGGTC
TTATAAAGAAGACTCGTGAGCTCTCTATTCTCTGTGACGCTCACATCGGTCTCATCGTCT
TCTCAGCCACCGGAAAGCTTTCGAGTTCTGCTCCGAACAGAACAGGATGCCTCAACTCA
TTGACCGATACTTGATACCAACCGATTGCGACTTCCTGATCATCATGACGACAGGAGC
AATTGCACCATGAGATGGAACCTACTAAGAAGAGAGACATGTAACCTTGAGCTTCGTCTGC
GTCCATTCCATGGACATGACTTAGCCTCCATTCTCTTAATGAGCTTGACGGACTCGAGA
GACAGCTAGAACATTCTGTCTCTCAAGTCCGTGAGCGTAAGAGGAGGATGCTAGAGAAG
ATAACAACAACATGTACCGTTGGCTTTCATGAGCATCGTGCAGCGATGGAGTTTCAACAAG
CTGGGATAGATACCAAACAGGGGAGTATCAACAGTTTATAGAGCAGCTTCAGTGCTATA
AACCAGGGGAGTATCAGCAGTTTCTAGAGCAGCAGCAACAACAACCAACAGCGTTCTTC
AGCTTGCTACACTTCTTCTGAGATTGATCCTACTTACAATCTCCAGCTTGCTCAGCCTA
ATCTTCAAACAGATCCAACGGCCCAAGATGATTAATACAATCTCAATAGATATCTACTC

TTTCTTTATGGAGACAGATTTCATGAACCTTTTATTACCTATATTTTGATAAGCCAGTGTCT
TCTTTTGTGTGGCTATGGAAACCTTGTTTAAAGCACAATGCACCTGAGTTCTTGTTTATA
TAATTAATCATCATTATTACATANNAAANAANNAAAAAAAAAAAAAA
>G156 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRGKIEIKKIENQTARQVTFKRRRTGLIKKTRELSILCDHIGLIVFSATGKLSEFCSE
QNRMPQLIDRYLHTNGLRLPDHDDQQLHHEMELLRRRETCLNLELRLRPFHGHDLASIPP
NELDGLERQLEHSVLKVRERKRRMLEEDNNMYRWLHEHRAAMEFQQAGIDTKPGEYQQF
IEQLQCYKPGYQQFLEQQQQPNVSLQLATLPSEIDPTYNLQLAQPNLQNDPTAQND*
>G1584 (160..1281)
ATTCACATTTTATTTATCTTTCCATTTAGCCATTCTGTTCCCTGTCTCTTCTCCTCTC
TTTTTGACACATCACATCATCATCATCATTCACATCAATCATCATCATATGCAT
ACACATACATCTGTGTTCTGCGGATCGAGTTAATTAGTTATGGCTTCTTCGAATAGACAC
TGGCCAAGCATGTTCAAGTCCAAACCTCATCCCCATCAATGGCAACATGACATCAACTCT
CCTCTCTTGCCTTCTGCTTCTCACCAGTCTTCTCCTTTCTCTTCAGGATGTGAAGTGGAG
AGGAGTCCAGAGCCAAACCAAGATGGAATCCAAAGCCAGAGCAGATTCCGATACTTGAA
GCAATCTTTAACTCCGGGATGGTGAATCCTCCAAGAGAGGAGATCAGGCTTCAAGAATAC
GGCCAAGTCGGTGATGCTAACGTCTTCTACTGGTTCCAAAACCGTAAGTCCCGTAGTAAA
CACAACTCCGCCTCCTCCACAACCCTCCAAACACTCTCTCCCTCAAACGCAACCGCAG
CCGCAGCCGCAACCTTCGGCTTCTCTCTCCTCTTCTCCTCTCTCCTCTCCTCCAAATCC
ACCAAACCCGAAAAAGCAAGAACAAGAACAACTAATCTCTCTTGGGTGGTAGTCAA
ATGATGGGGATGTTTCCACCGAACCGGCGTTTCTCTCCCGGTCTCCACTGTCCGAGGG
TTTGAAGGTATCACCGTCTCATCCCAATTAGGGTTTCTCTCCGGTGATATGATTGAGCAA
CAAAAACCGGCTCCAACGTGTACCGGACTCCTGTGAGTGAGATCATGAACGGTAGTGTG
AGTTATGGAATCATCATCAACAACACTTGAGTGAGAAAGAAGTTGAAGAAATGAGGATG
AAGATGTTGCAACAGCCACAGACTCAGATTTGTTACGCTACCACTAATCATCAAATAGCT
TCTTACAACAACAACAACAACAATAACATCATGCTTCATATCTCTCCCACTACTTCT
ACTGCCCACTATTACTACTTCTCGATTCTCTCGCTACTGTCCCATCAACTTCGGACCAG
CTTCAAGTTCGAAGCGGACGACGAATAAGAGTTTTCATCAATGAAATGGAGCTTGAAGTG
AGCTCAGGACCGTTCAATGTGAGGGATGCATTTGGGGAAGAGGTTGTTCTGATTAATTCC
GCGGGTCAGCCCATTTGTCACCGATGAATATGGCGTCGCTCTTACCCTCTTCAACACGGA
GCCTCGTACTATCTGATCTAGTCGTGTGGGAGATTGAGTTTGAAGAAGAAATTAAGACC
TGTCTCTTTCTTTCACCATCTCTCGTACGTAGGCTTAAATGTTAAGATTTTATAAGTAT
TGGTTTCAGTTACCTGTTGTGACGGTGTATTATGTATGAGTTTCGGACAACATTCACAAA
CTCTCTCGTTAAATTGTTGACCAATAATATATGATGTGTGTTTCATTATTATCTAAAAA
AA
>G1584 Amino Acid Sequence (domain in AA coordinates: TBD)
MASSNRHWPSMFKSPHPHQWHDINSPLLPSSASHRSSPFSSGCEVERSPKPRWNPKE
EQIRILEAIFNSGMVNPPEBIRLQYGVGDANVFYWFQNRKSRSKHLRLHLNHSKHS
LPQTQPPQPPQPSASSSSSSSSSSSKSTKPRKSKNNNTNLSLGSQMMGMFPPEPAFLF
PVSTVGGFEGITVSSQLGFLSGDMIEQKPAPTCTGLLLSEIMNGSVSYGTHHQHLSEK
EVEEMRMKMLQPPQTQICYATTNHQIASYNNNNNNNNIMLHIPPTTSTATTITTSLSLAT
VPSTSDQLQVQADARIRVFINEMBLEVSSGPFNVRFDAFGBEVVLINSAGQPIVTDEYGVA
LHPLQHGASYILI*
>G1587 (1..816)
ATGGGCTACATCTCCAACAACCTCATCAACTATTTGCCCTCTCTACTACTCAACCT
CCTCTTCTTCTCACCCTGTGATATTAACGGCAATGATCACCATCAGCTCATAACCGCA
TCATCAGGAGAACAGATATTGATGAACGGAAAAACAACATTCCTGCGGCGGCGACTTGT
AGATGGAATCCGACGCCAGAGCAGATCAGACGCTAGAAGAGCTTTACAGAAGCGGAACA
CGGACGCCGACGACGGAACAGATCCAACAGATAGCATCTAAGCTTCGTAAATATGGGAGA
ATCGAAGGGAAGAACGTTTCTATTGGTTTCAGAATCATAAGGCTAGAGAGAGACTAAAA
CGCCGCCGTCGTGAAGGTGGTCTATTATCAAACCACATAAAGACGTCAAGGATTATCA
TCAGGTGGTCTCGAGTTGATCAGACAAAGCTCTGCCCATCTTTTCCACACACAAACCGA
CCACAGCCACAGCATGAATTAGATCCTGCGAGTTACAATAAAGACAACAATGCTAATAAT
GAAGATCATGGGACGACTGAAGAATCTGATCAGAGGGCATCAGAGTTGGTAAATACGCC
ACATGGAGAAATCTTGTACTTGGTCGATAACTCAACAACCGGAAGAGATTAATATCGAC
GAAATGTCAACGGAGAGAAGAAGAAGACGAGGACACCGGACTTTAAATCTCTTCCG
GTTAGGGAGTACCAAGAGAAAACAGGCCGTTGATAGAGAAGACGAAAGCATGCAACTAC

TGTTACTACTACGAGTTCATGCCTCTGAAGAACTGA
>G1587 Amino Acid Sequence (conserved domain in AA coordinates: 61-121)
MGYISNNNLINYLPLSTTQPPLLLTHCDINGNDHHQLITASSGEHDIDERKNNIPAAATL
RWNPTPEQITLLEELYRSGTRTPTEQIQIASKLRKYGRIEGKNVFWFQNHKARERLK
RRRREGGALIKPHKDVKSSSGGHRVDQTKLCPSPHTNRPQPQHELDPASYNKDNANN
EDHGTTEESDQRASEVGKYATWRNLVWTSITQQPEEINIDENVNGEEETRDNRTLNLF
VREYQEKTRGLIEKTKACNYCYYYEFMPLKN*
>G1588 (1..2232)
ATGTACCATCCAAACATGTTTGAGAGCCATCATATGTTTCGATATGACCCCAAAGAGTACC
TCTGATAACGACTTGGGAATCACCGGTAGCCGAGAAGATGACTTTGAGACCAAGTCAGGT
ACCGAAGTCACTACTGAGAATCCTTCTGGTGAAGAGCTTCAAGATCCTAGCCAACGTCCC
AACAAAAAGAAGCGTTACCATCGCCACACGCAACGCCAAATTCAAGAGCTCGAATCATTC
TTTAAGGAATGTCCTCATCCAGATGATAAGCAACGAAAAGAGTTGAGCCGTGATCTCAAT
TTAGAGCCTCTTCAAGTTAAGTTTTGGTTCCAAAACAAACGCACACAGATGAAGGCACAA
AGTGAGAGGCATGAGAACCAGATTCTAAAGTCAGACAATGACAAGCTCAGAGCAGAGAAC
AATAGATACAAAGAAGCTTAAGCAATGCTACATGCCCTAACTGTGGCGGTCCAGCTGCT
ATTGGAGAAATGCTTTTTGACGAACAACATCTCAGGATCGAAAATGCTCGGCTCCGCGAA
GAGATTGATAGGATCTCTGCTATTGCTGCGAAATACGTTGGGAAGCCGTTAGGATCGTCT
TTCGCTCCACTAGCGATCCACGCGCCTTCTCGTTGCTTGATCTTGAAGTTGGAACTTT
GGGAACGACACAGGCTTTGTAGGAGAAATGTATGGAACAGGGGACATTTTGAGGTCAGTT
TCGATTCCTTCTGAGACTGATAAGCCTATAATCGTGGAGCTAGCGGTTGCAGCTATGGAG
GAACTCGTGAGAATGGCTCAAACCTGGAGATCCTTTATGGCTTTCAACCGATAATTCAGTC
GAGATTCTCAACGAAGAAGAGTATTTTCAGAACGTTTCCGAGAGGAATTGGACCAAAGCCA
TTAGGATTAAGATCAGAGGCGTCAAGACAATCTGCAGTTGTTATAATGAATCACATCAAT
CTCGTTGAGATTCTCATGGATGTGAATCAATGGTCTTGTGTTTTCTCTGGGATTGTGTCA
AGAGCCTTGACACTTGAAGTTCTTTCAACTGGAGTTGCTGGGAACTACAACGGTGCTTTA
CAAGTGATGACAGCTGAGTTTCAAGTTCCATCACCCCTAGTCCCAACGCGTGAGAATAC
TTTGTGAGATACTGCAAACAACACAGTGACGGCTCTTGGGCTGTGGTTGATGTCTCTTTG
GACAGCCTTAGACCAAGTACTCCAATCTTAAGAACTAGAAGAAGGCCTTCAGGTTGTCTG
ATTCAGAATTTGCCATATGGTTATTCTAAGGTTACATGGATAGAGCATATGGAGGTAGAT
GATAGATCAGTTCAACATGTATAAACCGTTGGTTTCAGTCCGGTTTAGCTTTTCGGTGCG
AAACGTTGGGTGGCTACACTCGAACGACAATGCGAGCGGCTTGCTAGCTCCATGGCCAGC
AACATTCTCTGGTGATCTTTCCGTGATAACGAGTCTGAAGGAAGGAAGATATGTTGAAG
CTAGCTGAGAGAATGGTTATGAGTTTCTGCAGTGGTGTGGCGCGTGCAGTGCACACGCT
TGGACAACAATGTGACAACAGGATCCGATGATGTTCCGGTCATGACCCGCAAGAGTATG
GATGATCCAGGAAGACCTCCGGGTATTGTTCTTAGTGCAGCTACTTCATTCTGGATCCCA
GTTGCTCCCAAACGTTTCTTCTCCTCCGTGACGAAAATTCAAGAAAAGAGTGGGAT
ATTCGTCAAAATGGAGGTATGGTTTCAGGAAATGGCTCATATAGCCAATGGTCATGAACCT
GGAACTGTGTCTCCTTGTCTCCGAGTCAATAGTGGAAGTTCAGGCCAGAGCAACATGTTG
ATTCATACAAGAGAGCTGTACAGATGCATCAGGATCGTATGTGATTTACGCGCCAGTGGAT
ATAGTGGCGATGAATGTGGTTCTAAGCGGTGGAGATCCTGATTACGTGGCGTTGTTGCCG
TCTGGTTTTGCTATTTTACCGGATGGTTCCGTTGGAGGAGAGATGGGAATCAGCATCAG
GAAATGGTTTCTACTACTTCTTCTGGGAGTTGTGGTGGTTTCGCTTTTAACCGTTGCGTTT
CAGATTCTTGTGACTCTGTTCTTACAGCTAACTCTCACTTGGCTCGGTGGCTACGGTT
AATAGTCTGATCAAAATGTACGGTGGAGAGGATTAAAGCTGCTGTTTCTTGTGATGTTGGA
GGAGGAGCGTAG
>G1588 Amino Acid Sequence (domain in AA coordinates: 66-124)
MYHPNMFESHMFDMPKSTSDNDLGITGSREDDFETKSGTEVTENPSGEELQDPSQRP
NKKKRYHRHTQRQIQELESFFKECPHPDDKQKELSRDLNLEPLQVFWFQNKRTQMQAQ
SERHENQILKSDNDKLAENNRYKEALSNAATCPNCGGPAAIGEMSFDQHLRIENARLRE
EIDRISAIKAYVGKPLGSSFAPLAIHAPSRSLDLEVGNFNGQTGFVGEYGTGDILRSV
SIPSETDKPIIVELAVAAMEELVRMAQTGDPWLWSTDNSVEILNEEYFRTFFRGIGPKP
LGLRSEASRQSAVIMNHINLVEILMDVNQWSCVFSGIVSRALTLEVLSTGVAGNYNGAL
QVMTAEFQVPSPLVPTRENYFVRYCKQHSQSWAVVDVSLSLRPSTPILRTRRRPSGCL
IQELPNGYSKVTWIEHMEVDDRSVHNMYKPLVQSGLAFGAKRWVATLERQ CERLASSMAS
NIPGDL SVITSP EGRKSM LKLAERMVMSFCSGVGASTAHAWTTMSTTGSDDV RVMT R KSM
DDPGRPPGIVLSAATSFWIPVAPKRVDFLRDENS RKEWDILSNGGMVQEMAH IANGHEP

GNCVSLLRVNSGNSQSNMLILQESCTDASGSYVIYAPVDIVAMNVVLSGGDPDYVALLP
SGFAILPDGSGVGGDGNQHQEMVSTSSGSCGGSLLTVAFQILVDSVP TAKLSLGSVATV
NSLIKCTVERIKAAVSCDVGGGA*

>G1589 (179..2221)

ACCAAACCTCACATAGCAATCACACACATCTCCACAAACACAGCTTGAGATGATCATGAAA
CACGTGCATCCTCAGATCTCTATCAATCCAGCTTGGTGAAAGAGGTCAAGAAATGAAAG
AGAATCAAAGAAAACGACGTCGTTTCATTCTGTGTAAACACTACTAATTATACATAGAT
GGCTGCTTACTTTACGGAAACCCACCGGAGATCTCTGCCGGATCCGACGGTGGTCTTCA
AACGTTGATCCTCGATCGAATCCAACACTACTTACGTTTCAGTACACCAACAAGACAACGACTC
GAACAACAACAACACAGCAACAATAGCAACAACAACAACACAAACACAAACACAAACAA
CAACAACAGTAGTTTTCGTTTCTCGATTCCCACGCGCGCGCAGCCAAACGCGAGCCAGCA
GTTCTGTCGGAATACCACTCTCAGGTCACGAAGCTGCTTCCATTACAGCCGCCGACAACAT
CTCCGTACTTTCACGTTTATCCTCCGCGCGTGCAGTACAGTCTCTACGGTAGCCACCAAGT
GGATCCCACTCACCAGCAAGCCGCGTGTGAGACGCCACGCGCGCAGCAAGGCCCTCTCTTT
AACCCTCTCGTCTCAACAGCAGCAGCAACAGCAACATCATCAACAACACCAGCCTATTCA
CGTCCGATTTCGGGATCGGATCAGCGAAGATATCCGGGTCGGGTCTGGCTCTACAGGATC
GGGGGTAACAAACGGTATAGCTAATCTTGTAGCTCCAAGTACTTGAAGGCAGCACAAAGA
GCTTCTTGACGAAGTAGTCAACGCTGATTCCGATGACATGAACGCTAAATCCCAACTATT
CTCATCGAAAAAGGGTAGTTGCGGAAATGATAAACCTGTCTGGAGAATCATCGGCCGCGC
TGGAGGAGAAGGTTCCGGTGGCGGAGCAGAAGCAGCCGGGAAACGTCGCGTGGAGCTAGG
CACGGCAGAGAGACAAGAAATACAGATGAAGAAAGCAAACTTAGTAACATGCTTCATGA
GGTGGAGCAGAGATATAGACAGTACCACCAGCAGATGCAGATGGTGATCTCTTCGTTCTGA
GCAAGCGGCAGGGATAGGATCAGCGAAGTCATACACGTCGCTAGCATTGAAAACCATATC
AAGACAGTTCCGTTGCTTGAAAGAGGCGATCGCTGGTCAGATAAAAGCGGCCAACAAAGAG
TCTTGGGGAGGAAGATTCAAGTGTCTGGTGTGGGAGGTTTGGGGGTCGAGGCTCAAGTT
CGTGGACCACCATTGAGACAGCAAAAGAGCTCTTCAACAACCTGGGAATGATTCAACATCC
TTCCAATAATGCTTGGAGACCTCAACGTTGGTCTCCAGAACGAGCCGTCCTAGTTCTCCG
TGCTTGGCTCTTCGAACACTTTCTTCATCCATACCCTAAGGATTTCGGACAAGCACATGCT
AGCTAAGCAAAACGAGACTCACTCGTAGCCAGGTGTCTGAACCTGGTTTATAAACGCGAGAGT
TCGGTTATGGAAACCAATGGTGGAGAGATGTACATGGAGGAAATGAAGGAGCAGGCAAA
GAACATGGGATCCATGGAAAAGACTCCTTTGGATCAAAGCAACGAAGATTCTGCTTCAAA
GTCAACAAGTAACCAAGAAAAGAGCCCAATGGCGGACACTAATTACCATATGAATCCCAA
TCACAACGGTGACCTAGAAGGCGTCACTGGAATGCAAGGATGCCCAAGAGACTAAGAAC
CAGCGACGAGACAATGATGCAGCCAATAAATGCGGATTTCAGCTCCAACGAGAAGCTCAC
GATGAAGATTCTAGAAGAACGGCAAGGGATAAGATCAGATGGTGGCTACCCCTTTCATGGG
TAATTCGGGCAATACCAATGGATGAGATGTCAAGATTTGATGTAGTCTCAGACCAGGA
GCTCATGGCGCAAGGTACTCAGGAAACAACAATGGCGTGTCCCTCAGCTTAGGTTTACC
TCATTGTGATAGCTTGTCTGTCACGGACCATCAGGGTTTCATGCAGACCCACCATGGGAT
TCCTATAGGGAGAAGAGTGAAAATAGGAGAAACAGAGGAATATGGACCCGCCACCATCAA
TGGTGGTAGCTCGACCACAACCGCACATTTCATCAGCGGCAGCTGCCGCGGCTTACAATGG
GATGAACATACAGAACCAGAAGAGATATGTGGCTCAGTTATTGCCCGACTTCGTTGCATA
AACCCTCTCTCTAGAAGGAGAAACCGAAACAGGTTATTATATACGTTTCTAGTTTTTAA
TTAGTATATAGTTTCTCATACCATTGAACCAAAACAAGAAACAAATTTAATTTTAGTCT
TTGGTTATATATGCGCCGACGGGCTACGTCAGGGCCCTGACGTAGC

>G1589 Amino Acid Sequence (conserved domain in AA coordinates:384-448)

MAAYFHGNPPEISAGSDGGLQTLILMNPTTYVQYTQQDNDNSNNNNNSNNNNNTNTNTN
NNNSSFVFLDSDHAPQPNASQQFVGIPLSGHEAASITAADNISVLHGYPYPRVQYSLYGS HQ
VDP THQQAACETPRAQQGLSLTLSSQQQQQQHHQHQPIHVGFSGHGEDIRVSGSGTG
SGVTNGIANLVSSKYLKAAQELLDEVVNADSDDMNAKSLFSSKKGSCGNDKPVGESSAG
AGGEGSGGGAEEAAGKRPVELGTAERQEIQMKAKLSNMLHEVEQRYRQYHQMQMVISSE
EQAAGIGSAKSYTSLALKTISRQFRCLKEIAGQIKAANKSLGEEDSVSGVGRFEGSRLK
FVDHHLRQORALQQLGMIQHPSNNAWRPQRGLPERAVSVLRAWLFEHFLHPYPKDSKHM
LAKQTGLTRSQVSNWF INARVRLWKPMVEEMYMEEMKEQAKNMGSMKTPLDQSNEDSAS
KSTSNQEKSPMADTNYHMNPNHNGDLEGV TGMQGC PKRLRTSDETMMPINADFSSNEKL
TMKILEERQGI RSDGGYPFMGNFGQYQMDMSRFDVVSQDELMAQRYSGNNNGVSLTLGL
PHCDLSSTDHGFQTHHGIPIGRRVKIGETEEYGPATINGGSSTTTAHSAAAAAAYN
GMNIQNQKRYVAQLLPDFA*

>G160 (38..784)

TCAAATTTGTCATTTGTTTATTCAAATTTTTGAGAAAATGGTGAGAAGTACCAAAGGTCG
TCAGAAAATAGAGATGAAAAAATGGAAAACGAAAGCAACCTTCAGGTTACTTTCTCAA
AAGAAGATTCGGTCTTTTCAAAAAAGCTAGTGAACTTTGACATTAAAGTGGTGCAGAGAT
TCTGTTGATTGTGTTCTCTCCTGGTGGGAAAGTGTCTTTTGGCCATCCAAGTGTTC
AGAATCTATTCATCGCTTTTTCGAATCCTAACCATAATTCTGCCATTGTCCATCATCAGAA
CAACAATCTCCAATTTGTTGAAACCCGTCGGGATAGAAATATCCAATATCTCAACAATAT
ACTCACTGAGGTGCTGGCAAACAGGAAAAGGAGAAACAGAAGAGAATGGTTTTGACCT
ATTGAAAGAATCCAGAGAACAGTAGGAACTGGTATGAAAAAGATGTGAAAGATCTCGA
CATGAATGAAACCAACCAGCTGATATCTGCTCTTCAAGATGTGAAAAAGAACTGGTAAG
AGAAATGTCTCAATATTCTCAAGTAAATGTTTCGAGAATTACTTTGGTCAAAGTTCTGG
CGTGATTGGTGGTGGTAAATGTTGGCATTGATCTTTTGGATCAAAGAAGAAATGCATTCAA
CTATAATCCAAACATGGTGTTCCTCAATCATAACACCACCAATGTTTGGATACAACAATGA
TGGAGTTCTCGTTCGGATATCCAACATGAACACATGTCAAGTTACAACCTCAACCAGAG
CTAGAGTCTGAAGCTAGAAGAACATCCTAATCAATATTGCGTTATTTGGCTATGGTTA
CTGTTAGGATTGTTCTTGTATTGTGAGACTTAAGTTTGTCTTTCTTTAATTTGTTTCA
GTTGGTTGGTTTTTCATTTTATTTCGTCGTTTGTCTTTGTTTGGATATTTTGTGTA
TCCCAGAATAAATTTATTTATCCTTTAAAAA

>G160 Amino Acid Sequence (domain in AA coordinates: 7-62)

MVRSTKGRQKIEKKMENESNLQVTFKRRFGLFKKASELCTLSGAEILLIVFSPGGKVF
SFGHPSVQELIHRFSNPNHNSAIVHHQNNNLQLVETRPDRNIQYLNILTEVLANQBEKEK
QKRMVLDLLKESREQVGNWYEKDVKDLDMNETNQLISALQDVKKLVREMSQYSQVNVVSQ
NYFGQSSGVIGGGNVGIDLFDQRRNAFNYPNPMVFPNHTPPMFGYNNDGVLVPIISNMNYM
SSYNFNQS*

>G1636 (19..666)

GAGTAATCATCAACGATTATGGCGTCAAGTCAGTGGACGAGGTCGGAGGATAAGATGTTT
GAGCAAGCTTTGGTCTTTTTCCTGAAGGATCTCCTAATCGGTGGGAGAGAATCGCTGAT
CAGCTTCATAAATCTGCTGGTGAAGTTAGGGAGCATTACGAGGTCCTTGGTTCATGATGTT
TTCGAGATTGATTCTGGTTCGAGTTGATGTCCCTGATTACATGGATGACTCGGCGGGCTGCG
GCGGCGGGTGGGATTCCGCTGGTCAGATCTCTTTGGGTCTAAACATGGCGAGAGTGAA
CGCAAAGAGGAACTCCTTGGACAGAGAACGAACACAAATGTTTCTGATCGGATTAAAG
AGATATGGTAAGGGAGATTGGAGGAGTATCTCGAGAAACGTTGTGGTGACGAGGACACCG
ACGCAAGTCGCGAGTCACGCTCAGAAGTATTTCTGAGACAGAACTCGGTGAAGAAGGAG
AGGAAAAGGTCGAGCATCCATGATATAACTACGGTTGATGCTACTTTGGCTATGCCTGGG
TCTAACATGGACTGGACTGGCCAAACCGGAGTCCTGTTACGGCGCCGAGCAGCAACAG
ATTATGTCTGAGTTCGGTCAGCAATTGAATCCTGGTCATTTCGAGGATTTTGGGTTTCGG
ATGTGATG

>G1636 Amino Acid Sequence (domain in AA coordinates: 100-165)

MASSQWTRSEDKMFEQALVLFPEGSPNRWERIADQLHKSAGEVREHYEVLVHDFEIDSG
RVDVPDYMDDSAAGWDSAGQISFGSKHGESERKRGTPTWENEHKLFLIGLKRYGKGD
WRSISRNVVTRTPTQVASHAQKYFLRQNSVKKERKRSSIHDITTVDATLAMPGSNMDWT
GQHGSVPVQAPQQQIMSEFGQQLNPGHFEDFGFRM*

>G1642 (1..1077)

ATGGGTCATCACTCATGCTGCAACAAGCAAAAGGTGAAGAGAGGGCTTTGGTCACCTGAA
GAAGACGAAAAGCTCATCAACTACATCAATTCATATGGCCATGGATGTTGGAGCTCTGTT
CCTAAACATGCAGGTTTGCAGAGATGTGGAAGAGTTGTAGATTAAAGATGGATAAATTAT
CTAAGACCTGATCTTAAACGTGGAAGCTTCTCTCCTCAAGAAGCTGCTCTTATCATTGAG
CTTCACAGCATTCTTGGTAACAGATGGGCTCAAATTGCTAAACATCTACCTGGAAGAACA
GATAACGAGGTCAAGAATTTCTGGAACCTCGAGCATTAAAAAGAAGCTCATGTCTCACCAT
CATCAGGTCATCATCATCATCTCTCTTCCATGGCGAGTTTGCTCACAAACCTTCCT
TATCAATGGATTCAACCTTACTACAGTCGACGATGAAAGTTCAAGATTCATGTCCAAT
ATCATCACAAACCTAACCCTAATTTTCATCACTCCAAGCCATCTCTCTCTCTCTCTCTCT
CATGTTATGACCCCATGATGTTCCCAACCTCTAGAGAAGGAGATTTCAAGTTTCTAACC
ACAAACAACCCAAACCAATCTCATCACCATGATAATAACCATTAACAACAACCTCGACATT
TTGTACCCACACCAACTATAACAATCATCATCAACCTTCACTTTCTTCTTGTCTCAT
GATAATAATCTCAATGGCCAGCGTTACCAGATTTCCAGCGAGTACCATTCTGGTTTC
CAAGAAACCTTCAAGATTATGATGATGCTAATAAACTCAACGTGTTTGTGACACCATTTC

AACGATAATGCCAAAAAGTTATTATGTGGAGAAGTTCTCGAAGGCAGAACTACTATCTTCC
TCCTCACCAATTTTCAACAAGATCACGGCCTTTTTCTTCCCACCACGTACAACCTTCAAATG
ACTTCTACGAGTGATCATCAACATCATCATCGAGTGGACTCATAATCAATCACATGATC
ATACCATCATCATCTCATCGTCGCCAATCTCTTGTGGACAGTACGTCATAACTTAA
>G1642 Amino Acid Sequence (domain in AA coordinates: TBD)
MGHHSCCNKQKVKRGLWSPEEDEKLINYINSYGHGCWSSVPKHAGLQRCGKSCRLRWINY
LRPDLKRGFSFPQEAALIELHSILGNRWAQIAKHLPGRTDNEVKNFWNSSIKKKLMSHH
HHGHHHHHLLSSMASLLTNLPYHNGFNPTTVDESSRFMSNIITNTNPNFITPSHLSLPSP
HVMTPLMFPTSRGDFKFLTNNPNQSHHHDNNHYNLDILSPTPTINNHHQPSLSSCPH
DNNLQWPALPDFPASTISGFQETLQDYDDANKLNVFVTPFNDNAKKLLCGEVLEGKVLSS
SSPISQDHGLFLPTTYNFQMTSTSDHQHHRVDSYINHMIIPSSSSSSPISCGQYVIT*

>G1747 (1..777)
ATGAAAATGATGCAAGAGGAGGGAACCGAAAAGGTCCATGGACAGAACAGGAAGACATA
CTTCTGGTAAATTTTGTTCACCTTATTTGGAGATCGACGATGGGATTTTATAGCAAAAAGTA
TCAGGTTTGAACAGAACAGGAAGAGTTGACGGCTAAGATGGGTTAATTACCTACATCCT
GGTCTCAAACCTGCAAGATGACGCCCTCAAGAAGAGCGCCTCGTCTTGAGCTTCACGCT
AAGTGGGGAACAGGTGGTGCAGAAATAGCCCCGAAAATTGCCGGGACGAAACGGATAACGAG
ATAAAGAACTACTGGAGGACTCATATGAGGAAGAAAGCTCAAGAAAAGAAGCGTCTCTGTT
TCCCCAACTTCTCTATTTTCCAACCTGCAGCTCGTCATCTGTGACCACTACCACCACCAAT
ACTCAAGATACATCGTGCCACTCGCGTAAATCTTCAGGGGAAGTGAGCTTTTACGACACT
GGAGGTTCCCGATCCACTAGAGAGATGAATCAAGAAAACGAAGACGTGTACTCGTTGGAT
GATATATGGAGAGAGATTGATCACTCAGCAGTAAACATAATAAAACCGGTTAAAGACATC
TACTCAGAACAAGCCATTGCTTAAGTTACCCAAATCTAGCTTCACCATCATGGGAAAGC
TCATTGGATTCTATATGGAACATGGATGCAGATAAAAGTAAGATATCGTCTTACTTTGCA
AATGATCAGTTTCTTTCTGTTTCCAACACAGTAGATCACCATGGTCGTCAGGTTAA

>G1747 Amino Acid Sequence (domain in AA coordinates: 11-114)
MKMMQEEGNRKGFWTEQEDILLVNFVHLFGDRRWDFIAKVSGLNRTGKSCRLRWVNYLHP
GLKRGKMTPOEERLVLELHAKWGNRWSKIARKLPGRTDNEIKNYWRTHMRKKAQEKRPV
SPTSSFSNCSSSSVTTTTNTQDTSCHSRKSSGEVSFYDTGGSRSSTREMNQENEDVYSLD
DIWREIDHSAVNIIPVKDIYSEQSHCLSYPNLASPSWESSLDSIWNMDADKSKISSYFA
NDQFPFCFQHSRSPWSSG*

>G1749 (59..535)
CAACACTTCTCAGTGACCGTGAGCAACGAATTATTTTCAGTTCAACGACTCCGCGGAAAT
GGAAAATTCAGAAAATGTTCCCTCTTACGATCAAAACATCAATTTCACTCCTAATTTGAC
GAGAGATCAAGAACATGTGATCATGGTCTCTGCTTTGCAACAAGTAATATCCAACGTCGG
AGGTGACACGAACCTCGAATGCATGGGAAGCTGATCTTCACCTTTGAACGCTGGCCCTTG
TCCTCTTTGTAGTGTACCGGCTGCTACGCTTGGCTCTTCCACGACACGAGGCGGATAAT
TAAGAAGGAGAAGAAGCACAAAGGAGTGAGGAAAAAACCATCAGGTAAATGGGCGGCGGA
GATATGGGATCCGAGTTTGAAGTAAGGAGATGGCTTGAACGTTTCCAACAGCGGAGAT
GGCGGCTAAGGCTTACAACGATGCGGCGGCTGAGTTTGTGCGGAAGAAGATCAGCAAGACG
TGGCACAAAGAACGGAGAGGAAGCATCTACCAAGAAGACGACTGAGAAAAATTACGGAG
AAGGAGCACGTATAGAAAGGCAGGAAGAGGCATCTTACTTGCTTCACAAGTAAATCAGAA
TTTTTTTTGAAAAGTAAAAACGTTATTTTGTGTTGTAATAAAATAAAGTAAAAACAAATAT
TGCTAACGCAAGACTTATCAAGTTCAGTCGTGACTGTGAGTGTGTTTTATGTATCTTAC
TTCATTTTTTGTCTTTCAATTGTGTGTGTGTGTGT

>G1749 Amino Acid Sequence (conserved domain in AA coordinates: 84-155)
MENSENVPSYDQININFTPNLTRDQEHVIMVSALQQVISNVGGDTNSNAWEADLPPLNAGP
CPLCSVTGCGYCVFPRHEAIKKKKHKGVRKKPSGKWAAEIWDPSLKVRRWLGTFTAE
MAAKAYNDAAAEFVGRRSARRGTKNGEEASTKKTTTEKN*

>G1751 (117..923)
AAACACAAACAAAACATATTTTCAATCTCCAGGTGCTTTACACCAACAGAGTCGCAAG
AAAACAAAACCAAACTCGGATTTAGTTTGACAGAAGAAGGAATCGAGAGTCGGGTATGC
ATTATCCTAACAACAGAACCGAATTCGTGCGAGCTCCAGCCCCAACCCGGTATCAAAGG
AGCAGTTGTACCGGAGCAAGAGCTTTCAGTTATTGTCTCTGCTTTGCAACACGTGATCT
CAGGGGAAAACGAAACGGCGCGGTGTCAGGGTTTTTCCAGTGACAGCACAGTGATAAGCG
CGGGAATGCCTCGGTTGGATTGACACACTGTCAAGTCTGTAGGATCGAAGGATGTCTCG
GCTGTAACACTTTTTTCGCGCCAAATCAGAGAATTGAAAAGAATCATCAACAAGAAGAAG

AGATTACTAGTAGTAGTAACAGAAGAAGAGAGAGCTCTCCCGTGGCGAAGAAAGCGGAAG
GTGGCGGGAAAAATCAGGAAGAGGAAGAACAAGAAGATGGTTACAGAGGAGTTAGGCAAA
GACCTTGGGGAAAAATTTGCAGCTGAGATCAGAGATCCTAAAAGAGCCACACGTGTTTGGC
TTGGTACTTTTCGAAACCGCCGAAGATGCGGCTCGAGCTTATGATCGAGCCGCGATTGGAT
TCCGTGGGCCAAGGGCTAAACTCAACTTCCCCTTTGTGGATTACACGTCTTCAGTTTCAT
CTCTGTTGCTGCTGATGATATAGGAGCAAAGGCAAGTGAAGCGCCAGTGTGAGCGCCA
CAGATTGAGTTGAAGCAGAGCAATGGAACGGAGGAGGAGGGGATTGCAATATGGAGGAGT
GGATGAATATGATGATGATGATGGATTTTGGGAATGGAGATTCTTCAGATTGAGGAAATA
CAATTGCTGATATGTTCCAGTGATAAATGAGCTCTTCTTGTGGCGTTTTTTGGAGTTA
AGTGCAAGAAGAGATTGACACTGTGGCTTGTAAAGTGAACAAGAACAAGAAAGCATGT
AATTAGTAGTCTCATTCTTTTGTGTTGTGGTCAATTCTATGTTTATCTCATATAAAATCTG
AGTTAAACCTATCTGAGGAGAGAGTAAATAAAGAGGTTAAGAA

>G1751 Amino Acid Sequence (domain in AA coordinates: TBD)
MHYPNNRTEFVGAPAPTRYQKEQLSPEQELSVIVSALQHVISGENETAPCQGFSSDSTVI
SAGMPRLSDSTCQVCRIEGCLGCNYFFAPNQRIEKNHQEETSSSNRRRESSPVAKKA
EGGGKIRKRKNKKNKYRGVRQRPWGKFAAEIRDPKRATRVWLGTFFETAEDAAARAYDRAAI
GFRGPRAKLNFPFVDYTSVSSPVAADDIGAKASASVSATDSVEAEQWNGGGGDCNME
EWMNMMMMMDFGNGDSSDSGNTIADMFAQ*

>G1752 (25..756)

AAAAAAAAAAAAAAAAAACTTATGGAATATTTCCCAATCTTCCATGTATTCATCTCCA
AGTTCTTGAGCTCATCACAAGAATCACTCTTATGGAACGAGAGCTGTTTCTTGGATCAA
TCATCTGAACCTCAAGCCTTCTTTTGCCTAATTATGATTACTCCGATGACTTTTTCTCA
TTTGAGTCACCGGAGATGATGATTAAGGAAGAAATCAAAACGGCGACGTTTCTAACTCC
GAAGAAGAAGAAAAGGTTGGAATTGATGAAGAAAGATCATACAGAGGAGTGAGGAAAAGG
CCGTGGGGGAAATTTGCAGCGGAGATAAGAGATTCAACGAGGAATGGAATTAGGGTTTGG
CTCGGGACATTTGACAAAAGCCGAGGAAGCCGCTCTTGCTTATGATCAAGCGGCTTTCCGC
ACAAAAGGATCTCTTGCAACACTTAATTTCCCGGTGGAAGTGGTTAGAGAGTCGCTAAAG
AAAATGGAGAATGTGAATCTTCATGATGGAGGATCTCCGTTATGGCCTTGAAGAGAAAA
CATTCTCTTCGAAACCGGCCTAGAGGGAAAAAGCGATCCTCTTCTTCTTCTTCTTCT
TCTAATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AAGCAGAGTGTTGTGAAGCAAGAAAGTGGTACACTTGTGGTTTTTGAAGATTTAGGTGCT
GAGTATTTAGAACAACCTTCTTATGAGCTCATGTTGATCTTGTAAATTGATTTCAGCAAAAG
CCACTATTAAACTTTAATTTTGTGATAATTAATCTTGAATTTGTTTTGTTTCTTCTGCA
ATTTCTTTGGTTCTCTTATTTTGTGTTGTGATCCAAATGAAATTATTGGAAGAGATG
GTGATGTTAAAGTGATATATATAAAAAAAAAA

>G1752 Amino Acid Sequence (domain in AA coordinates: TBD)
MEYSQSSMYSSPSSWSSSQESLLWNESCFLDQSSEPQAFPCPNYDYSDDFFSFESPEMMI
KEEIQNGDVSNSEEEKVGIDEERSYRGVRKRPWGKFAABIRDSTRNGIRVWLGTDFDKAE
EALAYDQAFAFKSLATLNFPEVVRESLKKMENVNLHDGSPVMALKRKHSLRNRPR
GKKRSSSSSSSSSSSSSSSSSTSSRSSSKQSVVKQESGTLVVFEFLGAEYLEQLLM
SSC*

>G1763 (33..977)

GTACATTTTTTTTTGTATTTTCAGGAACTCCGATGGCGGATCTCTTCGGTGGTGGCCACG
GCGCGAGCTTATGGAAGCACTTCAACCTTTTACAAAAGTGCTTCCACGTCTGCTTCAA
ATCCTGCGTTTGCCTCCTCAAACGATGCGTTTGGCTCTGCCCCAAACGACCCATTTCTT
CTTCTTCTTACTATAATCCTCATGCATCTTCTTCCCTTACATTCACAACTTACC
CGGATATTTTATTCTGGATCCATGACCTATCCATCTTCATTGCGGTGCGATCTTCAACAAC
CCGAAAACTACCAATCTCAGTTCCATTACCAAAACACTATCACTTACACTACCAAGACA
ACAACACTTGCATGCTCAACTTCATTGAGCCGAGCCAACCGGATTTATGACCCAACCGG
GTCCGAGTTGCGGTTCGGTTTCAAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGGAAAATGGGTGCGCGGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTCGACACGGCTGAAGAAGCCGCTTGGCTTATGATCGCGCCGCTTTAAGCTTC
GTGGTGACTCGGCTCGGCTTAACCTTCCAGCTCTCCGATACCAACCGGCTCGTCTCCGT
CTGACGTTGGCGAATACGGACCTATTCAAGCTGCCGTTGACGCCAAGCTAGAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCGCGGCTTCTTCACTGAGCAGCCGTCAGCGCCACAACAACATTCCGGGTGCGGTGAAA
GTGATGGGTGCGGTTACCGACTTCGGATGTTATGGTGAGGAGATGTGCCAAGAGCCAG

BNSDOCID: <WO 03013227A2 | >

GCGGCAAACGGAGCAATCTTGAAGCAGTAGACGGAGAGGCAAAGATCCACATCGTTGAC
ATAAGCTCCACGTTTGCACCTCAATGGCCGACTCTTCTAGAAGCTTTAGCCACAAGATCA
GACGACACGCCCTCACCTAAGGCTAACCACAGTTGTCGTGGCCAACAAGTTTGTCAACGAT
CAAACGGCGTCGCATCGGATGATGAAAGAGATCGGAAACCGAATGGAGAAATTCCGTAGG
CTTATGGGAGTTCTCTTCAAATTTAACATTATTCATCACGTTGGAGATTTATCTGAGTTT
GATCTCAACGAACTCGACGTTAAACCAGACGAAGTCTTGGCCATTAACTGCGTAGGCGCG
ATGCATGGGATCGCTTCACGTGGAAGCCCTAGAGACGCTGTGATATCGAGTTTCCGACGG
TTAAGACCGAGGATTGTGACGGTCGTAGAAGAAGAAGCTGATCTTGTGCGAGAAGAAGAA
GGTGGCTTTGATGATGAGTTCTTGAAGGGTTTGGAGAATGTTTACGATGGTTTAGGGTT
TGCTTCGAGTCATGGGAAGAGAGTTTCCAAGGACGAGCAACGAGAGGTTGATGCTAGAG
CGTCAGCGGGACGTGCGATCGTTGATCTTGTGGCTTGTGAGCCGTCGGATTCCACGGAG
AGGCGAGAGACAGCGAGGAAGTGGTCGAGGAGGATGAGGAATAGTGGGTTTGGAGCGGTG
GGGTATAGTGATGAGTGGCGGATGATGTCAGAGCTTTGTTGAGGAGATATAAAGAAGGT
GTTTGGTCGATGGTACAGTGTCTGATGCCGCCGAATATTCCTTTGTTGGAGAGATCAG
CCGGTGGTTTGGGCTAGTGCGTGGCGGCCAACGTAA

>G1767 Amino Acid Sequence (domain in AA coordinates: 255-272)

MDTLFRLVLSLQQQSDSIITNQSSLSRTSTTTTGGSPQTAYHYNFPQNDVVEECFNFFMD
EEDLSSSSHHNNHNNHNNPTYYSPFTTPTQYHPATSSPTSSATAAALASPYSSSHNN
DPSAFSIPQTPPSFDFSANAKWADSVLLEAARAFSDKDTARAQQILWTLNELSSPYGDTE
QKLASYFLQALFNRMGTSGGERCYRTMVTAATEKTCSESTRKTVLKFQEVSPWATFGHV
AANGAILEAVDGEAKIHVIDISSTFCTQWPTLLEALATRSDDTFLRLTTVVVANKFVND
QTASHRMMEKIGNRMEKFARLMGVPFKNIIHHVGDLSFDLNELDVKPDEVLAINCVGA
MHGIASRGSPRAVSISSFRRLRPRIVTVVEEADLVGEEEGGFDEFLRGFGECLRWFRV
CFESWEESFPRTSNERLMLERAAGRAIVDLVACEPSDSTERRETARKWSRRMRNSGFGAV
GYSDEVADDVRALLRRYKEGVWSMVQCPDAAGIFLCWRDQPVVWASAWRPT*

>G1778 (1..627)

ATGATGGGATACCAAACAACTCTAATTTCTCCATGTTTTTTTCTCGGAAATGACGAC
CAAAACCACCACAACCTACGATCCTTATAATAATTTCTCTTCATCAACTTCTGTTGATTGC
ACTCTCTCACTTGAACACCCCTCTACTCGTCTCGACGACCACCATAGATTTTCTCTGCT
AATTCTAACACATCTCCGGCGACTTTTATATTACGGAGGAAACGCTAAGACTTCTTCG
TACAAGAAGGGTGGTGTGCTCATAGCCTACCTCGCCGTTGTGCTAGCTGCGACACCACT
TCAACTCCTCTATGGAGAAACGGACCAAAGGACCTAAGTCGTTATGTAACGCGTGTGGA
ATCCGATTCAAGAAAGAGGAGAGGCGTGCGACGGCCAGAACTTAACGATCTCCGGTGGA
GGTTCATCAGCGGCAGAGTCCAGTAGAGAATTCGTACAACGGAGGTGGAACTATTAC
AGTCATCATCATCACTATGCCTCGTCGTCGCGTGTGGGCTCATCAGAACACACAA
AGAGTTCCATATTTCTCACCGGTTCCGGAGATGGAATATCCCTACGTGGATAACGTCACG
GCTTCTCTTTTATGTCTTGAATTGA

>G1778 Amino Acid Sequence (domain in AA coordinates: 94-119)

MMGYQTNSNFSMFSSENDQNHNYDPYNNFSSSTSVDCITLSLGPSTRLDHHRFSSA
NSNNISGDFYIHGNAKTSYKKGVAHSLPRRCASCDTTSTPLWRNGPKGPKSLCNACG
IRFKKEERRATARNLTISGGSSAAEVPVENSYNGGNYSHHHHHYASSSPSWAHQNTQ
RVVPYFSPVPEMEYPYVDNVTASSFMSWN*

>G1789 (108..413)

CAAGGACTCTGCGACATCTGTGCAACATATCATTTCTCAGAATCTCTTTCTTTCTAGG
TTTATTACTACACAAAACCAACATCATCAACTTTAGTTACTAAACAATGGCATCAGGCT
CAATGTCTTTCTTATGGCTCTGGCTCATGGACTGTTAAGCAGAACAAAGCCTTTGAGCGTG
CTCTAGCAGTCTATGACCAAGACACTCCGGACCGTTGGCACAATGTTGCTAGAGCTGTTG
GTGGTAAACACCAAGAAGCTAAGAGACAGTATGACCTTCTAGTTCGTGACATCGAAA
GCATCGAGAATGGTCACGTGCCATTCCCTGACTACAAGACTACTACAGGAAACAGCAACA
GAGGCAGGCTCGTGATGAGGAAAAGAGGATGAGAAGCATGAAGTGCAGTGAGACAAGA
AGCAACAAAACCTAACTACGTATGATCGTCAAAAATAAAGAGAATCACTTCAGAGAGATG
TGTTTTTTTCAATGTCTGACGAATCAATGTTTTTTTCTTGCAATTTCTCATGTTTTTCCC
TAAGAAATGGTTTTTTTTTCGAGGCAACAAAAA

>G1789 Amino Acid Sequence (domain in AA coordinates: 1-50)

MASGSMSSYSGSWTVKQKAFERLAVYDQDTPDRWHNVARAVGGKTPEEAKRQYDLLV
RDIESIENGHVFPDYKTTTGNNSNRGLRDEEKRMRSMKLG*

>G1790 (63..1346)

GAAAAAGACTTCACCTTTTTTTTTTACTAATTAATTAGTTTTTTTTTCTCCTTTCCAAAA
CAATGGAGAATTTTCGTCGACGAGAATGGTTTTGCTTCTCTAAACCAAAACATCTTCACAC
GTGATCAAGAACACATGAAAGAAGAGATTTTCCATTGGAAGTCGTCGACCAATCAAAAC
CTACAAGCTTTCTTCAAGATTTTCAACCATCTTGATCATGATCATCAGTTTGATCATCATC
ATCATCATGGCTCCTCATCTTCACATCCTTTGCTCAGCGTCCAAACTACGTCTTCTTGTA
TCAATAATGCTCCTTTTCGAGCATTGCTCTTACCAAGAAAACATGGTCGATTTCTATGAAA
CTAAACCAAAATTTGATGAATCATCATCATTTCCAAGCAGTGGAAAACTCATACTTCACTC
GTAATCATCATCATCATCAAGAGATCAATTTGGTCGATGAACATGATGATCCTATGGACT
TGGAGCAAAACAACATGATGATGATGAGGATGATCCCTTTTGATTACCTCCTACAGAGA
CTTTCAAACCTATGAACCTTCGTAATGCCAGATGAAATTTTCATGTGTTTCTGCAGATAATG
ATTTGTTATAGAGCAACGAGTTTCAACAAGACCAAAACCATTTCTTACACGAAAGTTGTCTT
CTTCTTCTTCATCATCATCATGGAAGAAACCAAAAAGTCAACCTTAGTCAAAGGACAAT
GGACTGCTGAAGAAGACAGGGTACTGATTCAACTCGTGGAGAAGTATGGATTGCGTAAAT
GGTCGCATATCGCTCAAGTGTTACCGGAAGAATCGGGAACCAATGTAGAGAGAGGTGGC
ATAACCATTTGAGACCTGACATTAAGAAAGAAACATGGAGTGAAGAAGAGGACAGAGTGT
TGATAGAATTTACAAAGAGATTGGAAACAAATGGGCAGAGATTGCGAAAAGACTCCCGG
GAAGAACAGAGAACTCGATCAAGAACCATTGGAACGCAACAAAAAGAAGACAATTCTCTA
AAAGAAAATGTAGATCTAAGTATCCAAGACCTTCTCTGTTGCAGGATTACATCAAGAGCT
TGAATATGGGAGCTTTGATGGCTTCTTCTGTTCTGCAAGAGGTAGACGCAGAGAGAGTA
ATAACAAGAAGAAGGATGTTGTTGTTGCGGTTGAGGAGAAGAAGAAGGAAGAGGAGGTGT
ATGGACAAGACAGGATGTGTCCTGAATGTGTGTTTACTGATGATTTTGGATTCAATGAGA
AGCTGCTTGAGGAAGGATGTAGCATTGACTCTTTGCTTGATGACATTCCTCAGCCTGACA
TTGATGCTTTTGTTCATGGGCTCTGATTTGTATTTTATTCTGCTTGTTCAGTTTGT
TGTTTTTTGTTTTGTCTTTTATACGAGACAGATTCCACCAAACTTCAATAATTTGAAAAG
ATATAAAATATTTTGTCTTTTAAAAAATAAAAAAAAAAAAAAAAAA

>G1790 Amino Acid Sequence (conserved domain in AA coordinates:217-316)

MENFVDENGFA SLNQNIFTRDQEHMKEDFPFEVVDQSKPTSFLQDFHHLDDHHDHDFHHH
HHGSSSSHP LLSVQTSSCINNAPFEHC SYQENMVDFYETKPNLMNHHHFQAVENS YFTR
NHHHHQEINLVDEHDDPMDLEQNNMMMRMIPFDYPPTETFKPMNFVMPDEISCVSADND
CYRATSFNKT KPF LTRKLS SSSSSSSSWKETKSTLVKGQWTAEDRVLIQLVEKYGLRKW
SHIAQVLPGRIGKQCRERWHNHLRPDIKKETWSEEDRVLIEFHKEIGNKWAEIAKRLPG
RTENSIKNHWNATKRRQFSKRKCRSKYPRPSLLQDYIKSLNMGALMASSVPARGRRRESN
NKKKDVVVAVEEKKKEEVYQDRIVPECVFTDDFGFNEKLLEEGCSIDSLDDIPQPD
DAFVHGL*

>G1791 (36..455)

ATGTACATGCAAAAACAAAACCTTAAAGCTTTTCATGGAACGTATAGAGTCTTATAACA
CGAATGAGATGAAATACAGAGGCGTACGAAAGCGTCCATGGGGAAAATATGCGGCGGAGA
TTCCGCACTCAGCTAGACACGGTGCTCGTGTTTGGCTTGGGACGTTTAAACACAGCGGAAG
ACGCGGCTCGGGCTTATGATAGAGCAGCTTTCCGCATGAGAGGCCAAAGGGCCATTCTCA
ATTTTCTCAGAGTATCAAATGATGAAGGACGGTCCAAATGGCAGCCACGAGAATGCAG
TGGCTTCCTCGTCTCGGGATATAGAGGAGGAGTGGTGGTGATGATGGGAGGGAAGTTA
TTGAGTTCGAGTATTTGGATGATAGTTTATTGGAGGAGCTTTTAGATTATGGTGAGAGAT
CTAACCAAGACAATTTGAACGACGCAAAACCGCTAGATCATCTACTTACTTACAGTGTA
ATGTTTTTGGAGTAAGAGTAATAATCAATATAATATACTTTAGTTTAGGAAAAA
AAAAA

>G1791 Amino Acid Sequence (domain in AA coordinates: TBD)

MERIESYNTNEMKYRGVRKRPWGKYAAEIRDSARHGARVWLGTFTNTAEDAARAYDRAAFG
MRGQRAILNFPHEYQMMKDGPNGSHENAVASSSSGYRGGGGDDGREVIEFEYLDSDLLE
ELLDYGERSNQDNCNDANR*

>G1793 (59..1783)

AGTGATTATTGATTAAACCCAAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTCTAACAACTGGCTTGGCTTTCTCTTTACCGAACAACCTTCTTTGCCTCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGGAACAACCTTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAACCGGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
CGACTCAGACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACACAGTAGCTATCATGAGCTTCAAGA

GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
 TGTAGACAAAGCTTCACCATCCGAGACCACGGGGATAACGCTAGCGGTGGAGCACTAGC
 CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
 CTATCGTGGTGTACAAGACATCGATGGACTGGTCGATATGAGGCTCATCTATGGGATAA
 TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
 CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACCTAAGTACTGGGGTCCTTC
 AACTACTACTAATTTCCCCATTACAACTACGAGAAAGAAGTAGAGGAAATGAAGCACAT
 GACGAGACAAGAGTTCGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
 TTCGATGTATCGAGGAGTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
 CCGAGTCGCCGGAAACAAAGACCTCTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
 AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
 GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCCATCGGAGGAGG
 CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
 GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
 CACCTCATCAAGACTTCAGCTTCAACCTTACCTCTAAGCATTCAACAACCATTAGAGCC
 TTTTCTATCTCTTCAACAATGACATCTCTCATTACAACAACAATGCTCAGGATTC
 CTCCTCTTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
 CAATTACTTGCAGCAACAGTCGAGCCAGAACCTCTCAGCAGCTCTACAATGCGTATCTTCA
 TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
 CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACACGGTATTGG
 TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAACAGATTA
 CGATATGCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTTGGACAGTGAGTCTGTTCA
 GGGGTCAAACCTGGTGGTGTCTTCACTATGTGGAATGAGTAAACAAGGATCTCTTCTT
 GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)

MNSNNWLGFPPLSPNNSSLPPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGGDEGEVVKV
 ADFLGVSKPDENQSNHLVAYNDSYFHTNSLMPVQSNVVAACDSNTPNNSSYHELQ
 ESAHNLQSLTSLMGTTAGNNVVDKASPSSETTGDNASGGALAVVETATPRRALDTFGQRTS
 IYRGVTRHRWTGRYEALWDNSCRREGQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGP
 STTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
 GRVAGNKDLYLGTFTSEEEAAEAYDIAAIKFRGLNAVTFEINRYDVKAILLESSTLPIGG
 GAAKRLKEAQALES SRKREAEMIALGSSSFQYGGGSSTGSGSTSSRLQLQPYPLSIQPPLE
 PFLSLQNDISHYNNNNAHDSSSFNHHYSIQTLHLHQQTNNYLQQSSQNSQQLYNAYL
 HSNPALHLHGLVSTSIVDNNNNNGSSGSYNTAAFLGNHIGIGSSSTVGSSTEEFPTVKTD
 YDMPSSDGTGGYSGWTSSESVQGSNPGGVFTMWNE*

>G1795 (27..422)

ACAAACACGCAAAAAGTCATTAATATATGGATCAAGGAGGTCGAGGTGTCGGTGCCGAGC
 ATGGAAAGTACCGGGGAGTTCGGAGACGACCTTGGGGAAAATATGCAGCAGAGATACGAG
 ATTTCGAGGAAGCACGGTGAACGTGTGTGGCTTGGAAACGTTTCGATACGGCAGAGGAAGCGG
 CTAGAGCCTATGACCAAGCTGCTTACTCCATGAGAGGCCAAGCAGCAATCCTTAACCTCC
 CTCATGAGTATAACATGGGGAGTGGTGTCTCTTCTTCCACCGCCATGGCTGGATCTTCCT
 CCGCCTCCGCTCCGCTTCTTCTTCTTAGGCAAGTTTGAATTGAGTACTTGGATG
 ATAGTGTTTTGGAGGAGCTCCTTGAGGAAGGAGAGAAACCTAACAGGGCAAGAAGAAAT
 GAGCGAGATATAATTATGATTATTTCTAA

>G1795 Amino Acid Sequence (domain in AA coordinates: 12-80)

MDQGGRGVGAEHGKYRGVRRRPWGKYAAEIRDSRKHGERVWLGTFDTAEAAARAYDQAA
 SMRQQAAILNFPHEYNMGSVSSSTAMAGSSSASASASSSRQVFEFEYLDLDSVLEELLE
 EGEKPNKGKKK* -

>G1800 (61..894)

CCATTATCATATCCTCTTCTTCTTCTTCACTATCAATCTTCTTCTTCCACTACAACACAA
 ATGGAGAAATCATCTCAATGAAACAATGGAAGAAGGGTCTGCTCGGGGTAAAGGCGGT
 CCACAAAACGCTCTTTGTCACTACCGTGGAGTCAGGCAAGGACTTGGGGCAAATGGGTG
 CTGAGATCAGAGAGGCCCAAGAGAGGGCAAGACTTTGGCTTGGCTCTTTCGCTACAGCT
 GAAGAAGCAGCTATGGCTTATGATGAGGCTGCCTTGAAACTCTATGGGCACGACGCATAC
 CTCAACTTACCTCATCTTCAAGGGAATACAAGACCTTCTCTGAGTAACTCTCAGAGGTC
 AAATGGGTACCTTCAAGGAAGTTTATATCTATGTTTCTTTCATGTGGTATGCTAAACGTG
 AATGCTCAGCCTAGTGTTCACATAATCCAGCAAAGACTAGAAGAACTCAAGAAAACCTGGA

CTTTTATCTCAATCCTATTCTTCTAGTTCTTCTCCACCGAATCAAAAACTAATACTAGC
TTTCTTGATGAGAAGACCAGCAAGGGAGAAAACAGACAATATGTTTCAAGGTGGTGATCAG
AAGAAACCAGAGATCGACCTGACCGAGTTTCTTTCAGCAACTAGGAATCTTGAAGGATGAA
AATGAAGCAGAACCAAGTGAGGTAGCAGAGTGTCAATCCCCCTCCACCATGGAACGAGCAA
GAAGAACTTGAAGTCCTTTTCAGAACTGAGAATTTAGCTGGGATACCCTGATCGAGATG
CCAAGAAGTGAAACCACAACATATGCAATTTGACTCCAGCAACTTCGGAAGCTATGATTTT
GAGGATGATGTATCCTTCCCTTCCATCTGGGACTACTACGGAAGCTTAGATTGAGTAAAA
GCAATTTAAGGTAGATCAAGATTGAGAAGTACACAAATGGTTTTGGATTTAGTGTAGCGT
TTTGAAAAAGAGACATAGGTAGTGAGAGTGCACTCTTTTATTATGCAGCAATAAAGTGAG
TCAGTGTAACACCGAGTTGTTTCGCTTTTTTTGGTATATTAATGAAGCATGTTTATTTTT
CGCTAAAAAAAAAAAAAAAAAAAAA

>G1800 Amino Acid Sequence (domain in AA coordinates: TBD)
MEKSSSMKQWKKGPARGKGGPQNALCQYRGVVRQRTWGWVAEIREPKKRRLWLGSFATA
EEAAMAYDEAALKLYGHDAYLNLPHLQRNTRPSLSNSQRFKWPSPRKFISMFPSCGMLNV
NAQPSVHI IQQRLEELKKTGLLSQSYSSSSSSSTESKNTSFLDEKTSKGETDNMFEGGDQ
KKPEIDLTEFLQQLGILKDENEAEPSVVAECHSPPPWNEQEETGSPFRTENFSWDTLIEM
PRSETTTMQFDSSNFGSYDFEDDVSPFSIWDYYGSLD*

>G1806 (1..1356)
ATGCAGAGCAGCTTCAAAACCGTTCCCTTCACTCCTGATTTCTACTCTCAATCCTCTTAC
TTCTTCAGAGGAGATAGTTGTCTTGAGGAGTTTCATCAACCAGTCAATGGTTTTTACCAT
GAAGAAGCTATCGATTTAAGTCCAAATGTCACTATTGCTTCAGCTAACTTACACTACACG
ACGTTTGATACGGTTATGGATTGTGGTGGTGGTGGTGGTGGCTTGAGGGAGAGACTTGAA
GGAGGAGAAGAGGAGTGTTTGGACACAGGGCAATTAGTGTAACAGAAAGGGACAAGATTA
GTAGGAGGAGGAGTAGGAGAAGTGAACAGCAGTTGGTGTGATTTCGGTTTCAGCTATGGCT
GATAACAGTCAACATACTGACACTTCCACAGATATTGATACTGATGACAAGACTCAGTTG
AATGGAGGTATCAAGGGATGCTATTGGCTACAAATTTTTCAGATCAATCCAATGTGAAA
TCTAGTGATCAAAGGACACTTCGTCGACTTGCTCAGAACCGGGAGGCTGCTAGGAAAAGT
CGGTTGAGGAAAAAGGCCTATGTTTCAGCAACTTGAGAATAGTCGAATCAGGCTTGACACAG
CTAGAGGAAGAGCTCAAAAGAGCTCGCCAACAGGGATCTTTGGTTGAAAGAGGAGTTTCA
GCGGATCACACGCATTTGGCAGCAGGAAATGGTGTCTTTTCATTTGAATTGGAATATACA
CGTTGGAAGGAGGAACATCAAAGAATGATCAACGACTTAAGATCGGGTGTGAATTCGCAG
TTAGGTGACAACGATCTACGCGTTCTAGTGGATGCTGTGATGAGTCACTATGATGAAATA
TTCAGGCTAAAGGGAATTGGCACTAAAGTTGAAGTCTTTCATATGCTCTCAGGCATGTGG
AAGACACCTGCCGAGAGATTTTTTCATGTGGTTAGGTGGATTTAGATCATCAGAGTTACTT
AAGATATTGGGGAACCATGTGGATCCATTGACGGACCAGCAGTTGATAGGCATTTGCAAC
CTTCAGCAATCGTCTCAACAAGCAGAGGATGCATTGTCAAGGCATGGAAGCTCTACAA
CAATCACTTCTCGAGACGCTTTCTTCTGCTTCTATGGGTCCAAACTCTTCAGCAAAATGTT
GCAGATTATATGGGTATATGGCTATGGCTATGGGCAAACTTGGCACTCTTGAAAACCTTC
CTTCGCCAGGCTGATTTATTGAGGCAACAACTCTGCAACAGCTTCACAGAATTCTCACC
ACACGACAAGCTGCTCGCGCCTTTTTGGTTCATCCACGATTATATTCTCGGCTTAGAGCA
CTTAGCTCTCTATGGTTAGCCAGACCTAGAGACTAA

>G1806 Amino Acid Sequence (domain in AA coordinates 165-225)
MQSSFKTVPTPDFYSQSSYFFRGDSCLEEFHQPVNGFHHHEAIDLSPNVTIASANLHYT
TFDVTMDCGGGGGGLRERLEGGEELDTGQLVYQKGTSLVGGGVGEVNSSWCDSVSAMA
DNSQHTDSTIDTDDKTQLNGGHQGMLLATNCSQSNVKSDDQRTLRRLAQNREARKS
RLRKAYVQQLENSRIQLAQLEELKRARQQGSLVERGVSAHHTLAAGNGVFSFELEYT
RWKEEHQRMINDLRSGVNSQLGDNDRVLVDVMSHYDEIFRLKGIGTKVEVFHMLSGMW
KTPAERFFMWLGGFRSSELLKILGNHVDPLTDQQLIGICNLQSSQQAEDALSQGMALQ
QSLLETLSASMGPNSSANVADYMGHMAMAMGKLTLENFLRQADLLRQOTLQQLHRIILT
TRQAARAFVVIHDYISRLRALSSLWLARPRD*

>G1811 (93..827)
AAAGGAGCATTGGTATCTCAACAATATTTGCCCTTTCTCTATCTCTCTCATCACTAT
TTGCCATCTCTTTCTCTCTCCCTCTCTTTCAAATGTCAATAAACCAATACTCAAGCGATT
TCCACTACCATTTCTCTCATGTGGCAACAACAGCAGCAACAACAACACCAAACGACG
TCGTGGAAGAAAAAGAAGCTTTTTTCGAGAAACCCCTTAACCCCAAGTGACGTCGGAAAAAC
TCAACCGCCTCGTCATCCCAAAACAGCAGCCGAGAGATACTTCCCACTAGCGGCCGCCG
CCGCAGACGCCGTGGAGAAAGGACTTCTCCTCTGCTTTGAGGACGAGGAAGGTAAACCAT

GGAGATTGAGATACTCGTACTGGAACAGTAGCCAGAGTTATGTCTTGACCAAAGGCTGGA
GCAGATACGTCAAGGAGAAGCACCTTGACGCCGAGACGTCGTTCTCTCCATCGACACC
GTTGAGACGGCGGAAGATTCTTCAATTGGCTGGAGAAGACGCGGTGACTCTTCTTCTCCT
CCGACTCTTATCGCCATGTTCATCAATCCAATGCCTCGCTCCAATATTATCCTCATGCAGGGG
CTCAAGCGGTGGAGAGCCAAAGAGGCAACTCGAAGACATTAAGACTGTTGCGAGTGAACA
TGGAGTGCCAGCTAGATTGCGACTGGTCCGAGCCATCCACACCTGACGGTCTTAACACAT
ATACAACCAATCACGACCAGTTTCAATTTCTACCCTCAACAACAACACTATCCTCCTCCGT
ACTACATGGACATAAGTTTACAGGAGATATGAACCGGACGAGCTAGAAGCCCACAAGGA
TTAAAAAAAAGCTTCACATCTGGTCTGTATGTGTCATAGATGTTGATTCTCTTAATTT
TACACAAGCTTCATTTTGCAATTATTTAAAGTAAAATCGTATTTTGATTCTCTTTAAATC
TCTCTCAATTTTCACTCTCTCTCTTTTCTTCTTATGTATTAGATTCTTTTACATAGCTA
ACACTTGTATAGAGAATTCAAAGTTCTGGCTATTTTCGAAAGTTATCTTTTCTCTTAAAA
AAAAAAA

>G1811 Amino Acid Sequence (domain in AA coordinates: TBD)

MSINQYSSDFHYHSLMWQQQQQQQHNDVVEEKEALFEKPLTPSDVGKLNRLVIPKQHA
ERYFPLAAAAADAVEKGLLLCFEDEEGKPRFRYSYWNSSQSIVLTGWSRYVKEKHLDA
GDVVLFRHRSDGGRFFIGWRRRGDSSSSSDSYRHVQSNASLQYYPHAGAQAQVESQRGNS
KTLRLFGVNMECQLDSWSEPSPTDGSNTYTTNHDQFHFYPQQQHYPPPYMDISFTGDM
NRTS*

>G182 (74..1366)

CGTCGACGATCAGATTCTTGCGTATAGCTGTATATATACACCAAGATACACTCATCATCG
TCATATATAGATTATGTGACGCGTCTCTGAGCTTCTTGACATGGAAAACCTTCAAGGAGA
CTTAACCGACGTCGTACGAGGAATCGGAGGCCACGTGTTATCACCGGAGACTCCTCCCTC
GAACATCTGGCCTCTTCTCTGTACATCCAACACCATCACCGTCAGATCTTAACATAAA
CCCCCTCGGAGATCCCTTTGTGAGCATGGACGATCCACTCCTCAAGAATAAACTCCAT
CACAAACTCCGGCTATTTCTCCACCGTAGGAGATAACAACAACAACATTACACAACAACA
TGGTTTCTTGGTTCCAAAGGTATTTGAGGAGGATCATATAAAGAGTCAATGTAGTATCTT
CCCAAGAATCCGGATCTCGCATAGTAACATCATCCACGATTCTTCTCCGTGTAATCTCC
GGCCATGTGCGGCTACGTTGTGCGAGCCGACGAGCCGCCCTCGCCGAGAGGCATCATCAA
CGTAGACACAAACAGTCTAGAACTGTCTATTGGTTGATGGTACCACGTTCTCTCTCGCA
GATTGAGATATCTTCCCTCGGAATCTAGGCCTTAAAGAAGGAAGAGTCAGGCAAGAA
GGTGGTGTGTATTCCGGCCCCGGCTGCAATGAACAGCCGATCAAGCGGAGAAGTGGTTCC
ATCGGATCTATGGGCTTGGCGTAAATACGGTCAAAAACCTATCAAAGGCTCTCCTTTTCC
AAGGGGTTATTATAGATGCAGCAGCTCAAAGGTTGTTGAGCAAGAAAGCAAGTCGAAAG
AAGCCGAACCGATCCAAACATGTTGGTGATTACATATACCTCCGAACATAACCATCCTTG
GCCCATCCAACGCAACGCTCTCGCCGGCTCCACACGCTCCTCCACCTCCTCCTCATCTAA
CCCTAATCCTTCCAAACCTCAACCGCAAACGTAACTCCTCATCCATTGGCTCCCAAAA
CACCATCTACTTGCTTCTCTCCACCACTCCTCCTCCTACCTCTCATCCTCCGCCATCAA
AGATGAACGAGGGGACGATATGGAGTTGGAAAACGTAGATGATGATGATAACCAGAT
TGCTCCATACAGACCGGAGCTTCATGATCATCAGCACCAACCAGATGATTCTTTGCAGA
TCTTGAAGAGCTAGAAGGAGATTCTTAAGCATGTTGCTTTCTCATGGCTGTGGCGGCGA
CGGGAAGGATAAAACGACCGCGTCCGATGGGATCAGCAATTTCTTCGGGTGGTGGGAGA
TAATAATTATAATAATTACGACGACCAAGACTCAAGGTCGTTATAGTATAGTGTAAATTA
CAGGTAACAAATTATATTAATAAGTTGAGCTTGTGAAAATGAAGATCATATGGTCTG
GTCAGGTTGGGGG

>G182 Amino Acid Sequence (conserved domain in AA coordinates: 217-276)

MCSVSELLDMENFQGLDLDVVRGIGGHVLSPETPPSNIWPLPLSHPTSPSDLNINPFGD
PFVSMDDPLLQELNSITNSGYFSTVGDNNNNIHNNGFLVPKVFEEDHIKSQCSIFPRIR
ISHSNIHDSPPCNSPAMSAHVAAAAASPRGIINVDTNSPRNCLLDVDTTFSSQIQIS
SPRNLGLKRRKSQAKKVVCIAPAPAMNSRSSGEVVPDLWAWRKYGQKPIKGSPPFRGYY
RCSSSKGCSARKQVERSRTDPNMLVITYTSEHNHPWPIQRNALAGSTRSSSTSSSSNPNS
KPSTANVNSSSIGSQNTIYLPSTTTPPTLSSSAIKDERGDDMELENVDDDDNQIAPYR
PELHDHQHPDDFFADLELEGDSLMLLSHGCGDGDKDTTASDGISNFFGWSGDNVYN
NYDDQDSRSL*

>G1835 (1..969)

ATGATTGGAACAAGCTTCCCCGAGGATCTTGATTGTGGCAACTTCTTTGACAACATGGAT
GATCTCATGACTTTCCCGGTGGAGATATCGATGTGCGTTTCGGCATAGGTGACTCCGAC

TCTTTCCCTACCATCTGGACCACTCATCACGACACGTGGCCTGCCGCTTCTGATCCTCTC
TTCTCTTCCAACACCAACTCTGATTCATCACCTGAGCTCTATGTTCCGTTTGAGGACATT
GTTAAGGTGGAAAGACCTCCAAGCTTTGTAGAGGAAACATTGGTTGAGAAGAAGGAAGAT
TCGTTTTTCGACAAACACTGATTCATCATCTTCTCATAGCCAATTGAGGAGCTCAAGTCCA
GTGTCGGTTCTCGAGAGCAGCTCCTCCTCGTCTCAAACCACCAACACAACCTCCCTTGT
CTCCCTGGAAAGCAGGTCCTCCACGCACAAAACGCCCTCGTCCACCTGTCCAGGATAAA
GATAGAGTCAAAGACAATGTGTGCGGTGGTGACTCGCGCTCATCATTTAGAATACCGAAA
CAGTTTCTCTCTGATCACAACAAGATGATCAACAAGAAGAAGAAGAAGGCCAAGATT
ACTTCTTCTCTTCTTCTCGTCCGGGATTGATCTTGAAGTCAATGGAACAACGTCGATTCTG
TATTTCTTCAGAGCAATATCCGCTTAGGAAATGTATGCACTGTGAGGTACCAAGACTCCA
CAGTGGAGGCTTGGTCCAATGGGTCCAAAGACACTTTGCAATGCGTGGGTGTACGTTAC
AAATCAGGGAGGCTTTTCCCGGAGTACCGTCCAGCTGCTAGTCCAACATTTACTCCAGCT
CTTCACTCAAACCTCACACAAGAAAGTGGCTGAAATGAGAAACAAGAGATGCAGTGATGGT
AGCTACATAACCGAAGAGAATGATCTGCAAGGGCTGATTCCGAACAATGCCTACATTGGC
GTAGACTAA

>G1835 Amino Acid Sequence (domain in AA coordinates: 224-296)
MIGTSFPEDLDCCNFFDNMDLMDPFGGDI DVGFGIGDSDFSPTIWTTHHDTWPAASDPL
FSSNTNSDSSPELYVPFEDIVKVERPPSFVEETLVEKKEDSFSTNTDSSSSHSQFRSSSP
VSVLESSSSSSQTNTTSLVLPKGHRPRTRPRPPVQDKDRVKNVCGGDSRLIIRIPK
QFLSDHNKMKKKKKKAKITSSSSSGIDLEVNGNNVDSYSSEQYPLRCKMHCEVTKTP
QWRLGPMGPKTLCNACGVRYKSGRLFPYRPAASPTFTPALHSNSHKVEMRNKRCSDG
SYITEENDLQGLIPNNAYIGVD*

>G1836 (47..610)
ATAACAAGCCTGAACACTAGAACTTCAAAAAAGAAAAAATCTTATGGAGAACAACAA
CGGCAACAACAGCTGCCACCGAAAGGTAACGAGCAACTGAAGAGTTTCTGGTCAAAGA
GATGGAAGGTAAGTATGATTTCAAAAATCACGACCTTCTATAACTCGTATCAAGAAGAT
TATGAAGTATGATCCGGATGTGACTATGATAGCTAGTGAGGCTCCAATCCTCTCTCGAA
AGCATGTGAGATGTTATCATGGATCTCACGATGCGTTCGTGGCTCCATGCTCAGGAAAG
CAAACGAGTCACGCTACAGAAATCTAATGTGCGATGCCGCGAGTGGCTCAAACCTGTTATCTT
TGATTTCTTGCTTGATGATGACATTGAGGTAAAGAGAGAGTCTGTTGCCGCCGCTGCTGA
TCCTGTGGCCATGCCACCTATGACGATGGAGAGCTGCCCTCAGGAATGGTAATTGGAAC
TCCTGTTTGTGTAGTCTTGGAATCCACCAACCACAACCACAATGCAGGCATGGCCTGG
AGCTTGGACCTCGGTGTCTGGTGAGGAGGAAGAAGCGCGTGGGAAAAAAGGAGGTGACGA
CGGAAACTAATAAGTGAATACGTTTTAGGGTATTTTCAAGGAATATGTAGTAAATAGT
CATGGATC

>G1836 Amino Acid Sequence (domain in AA coordinates: 30-164)
MENNNNGNNQLPPKNEQLKSFWSKEMEGNLDKFNHDLPIRIKKIMKYDPDVTMIASEAP
ILLSKACEMFIMDLTMRSLHAQESKRVTLQKSNVDAVAQTVIDFLDDDDIEVKRESV
AAAADPVAMPPIIDGELPPGMVIGTPVCCSLGIHQPPQMQAWPGAWTSVSGEEEEARGK
KGGDDGN*

>G1838 (132..1628)
TTCCCTGGCATTCTCTTTAGAACTTTCTGTACAAAATGCAAAACCTGAACCTCTAAAGCTA
AAAAAAAAGATTAGAGACTGTAAGTCTTTTATCAGATTTTCAACTAGGAAAAAAGTTAC
AATCTTTTTTGTATGGCTCCTCAATGACGAATTGCTTAACGTTTTCTCTGTACCAATGG
AGATGTTGAAATCAACTGATCAGTCTCACTTCTCTTCTTCTTACGACGATCTCTCTACTC
CTTATCTCATCGATAACTTCTATGCTTTCAAAGAAGAAGCTGAGATAGAAGCTGCTGCTG
CTTCAATGGCGGATTCAACAACCTTATCTACTTTTTTTCGATCATCTCTCAGACTCAGATTC
CAAAGCTGGAAGATTTCTCTCGGTGATTCTCTTGTCCGTTACTCTGATAACCAACAGAGA
CCCAAGACTCTTCTCTCTCACTCCATTCTACGATCCACGTCACCGCACCGTTGCCGAAG
GAGTTACAGGGTTCTTCTCTGATCATCATCAGCCAGATTCAAGACGATAAACTCGGGAC
CAGAAATCTTCGATGACTCAACAACCTTCCAACATCGGTGGTACTCATCTCTCCAGTCAG
TGCTGGAGTCATCAACGACGCGAAGTTAGGGTTTAACGGTGATTGCACCACCACCGGAG
GAGTTTTGTCTCTAGGGGTTAACAACACATCAGATCAACCTTTGAGCTGTAACAATGGCG
AGAGAGGTGGAACAGTAACAAGAAGAAAACAGTTTCTAAGAAGGAACATCAGATGATT
CAAAGAAGAAGATTGTGCAAAACATTGGGACAAAGAACCTCAATTTATCGTGGAGTCACCC
GACATAGATGGACTGGAAGATACGAAGCGCATCTATGGGATAACAGCTGTAGGAGGGAAG
GTCAAGCCAGAAAAGGACGTCAAGTGTACTTAGGTGGATATGACAAGGAAGATAGAGCAG

CTAGAGCCTATGACTTGGCAGCTTTAAATACTGGGGTTCTACTGCTACTACAAATTTTC
CGGTCTCGAGTTATTCAAAAGAACTTGAGGAAATGAATCACATGACCAAGCAAGAGTTTA
TTGCATCTCTTAGGAGGAAAAGTAGCGGTTTTTCGAGAGGAGCTTCAATATATAGAGGTG
TCACAAGGCATCATCAACAAGGTCGCTGGCAAGCAAGAATCGGCCGTGTCGCAGGAAACA
AAGATCTTTACCTCGGAACCTTTGCAACCGAAGAGGAAGCAGCAGAGGCTTATGACATTG
CAGCCATAAAGTTCAGAGGAATCAACGCAGTAACCTTTGAGATGAACAGGTATGACA
TTGAAGCTGTCAATAGATTCTTTACCTGTAGGAGGAGCAGCTGCGAAACGCCACAAAC
TCAAACCTCGCTCTGAATCTCCTTCTTCATCATCCTCTGACCATAACCTCCAACAACAAC
AGTTGCTTCCGTCCTCTTCTCCCTCGGATCAAAACCTTAACCTCAATCCCATGTGGCATT
CATTGAGCCTTCAGTTCTCTATTACCACCAGAACTTCTTTCAGCATTATCCTTTGGTCT
CTGACTCTACAATTCAAGCTCCTATGAACCAAGCTGAGTTTCTTGTGGCCTAACCACT
CTTACTAAATCATTGGTTCTGTTCTTGTCTTAGACTTCTATTACCGCACTAACCGATGAC
CCGAGGCTTATCTTCTTGATTCTGGCTATAAGGATGAATCTTCAAGTTCCTTTTAAAC
TGTAGGCTAAGACAGAAGTAGAGGGGAGAAAAGTTGAAGAATCTGAAACTTTTGGGGTCA
ATTTTGTATTAAATGTTTTCTTTGTCAAGGGTGGATTATCGGTTTTATTACTTATTTTT
TGAATGTAATCGGCCTATAACGGTATAACTCTGTTTCCATTATGAATATTTTCTCAA
TTGAAAAAAAAAAAAAAAAA

>G1838 Amino Acid Sequence (conserved domain in AA coordinates: 229-305, 330-400)

MAPPMTNCLTFLSPMEMLKSTDQSHFSSSYDDSTPYLIDNFYAFKEEAEIEAAAASMA
DSTTLSTFFDHSQTQIPKLEDFLGDSEFVRYSDNQETQDSSSLTFPYDPRHRTVAEGVTG
FFSDHHQPDFKTINSQPEIFDDSTTSNIGGTHLSSHVESSTAKLGFNGDCTTTGGVLS
LGVNNTSDQPLSCNNGERGGSNSKKKTVSKKETSDDSKKIVETLQRTSIRGVTRHRW
TGRYEHLWDNSCRREGQARKGRQVYLGGYDKEDRAARAYDLAALKYWGSTATTNFPVSS
YSKELEEMNHMTKQEFIASLRKSSGFSRGASIRGVTRHHQQGRWQARIGRVAGNKDLY
LGTFAEEEAAYDIAAIAKFRGINAVTNFEMNRYDIEAVMNSSLPVGGAAAKRHLKLA
LESPSSSSSDHNLQQQLLPSSSPSDQNPNSIPCGIPFEPVLYHQNFQHYPLVSDST
IQAPMNQAEFFLWPNQSY*

>G1843 (51..653)

CAGACATCACAATCAAATTAGGTCAGAAGAATTAGTCGGAGAAAACAGCCATGGGAAGAA
GAAAAGTAGAGATCAACGAATTGAGAACAAAAGCTCTCGACAAGTTACTTTCTGTAAAC
GACGAAATGGTCTCATFGAGAGAAAGCTCGTCAACTCTCAATCTTTGTGAATCCTCCGTCG
CTCTTATCATCATCTCTGCCACCGGAAGACTCTACAGCTTCTCTCAGGTGATAGCATGG
CCAAGATCCTCAGTCGTTATGAATTAGAACAGGCTGATGATCTTAAACCTTGGATCTAG
AAGAAAAAATCTTAATTATCTTTGCGACAAGGAGTTGCTAGAAACAATCCAATGCAAGA
TTGAAGAAGCGAAAAGCGATAATGTAAGTATAGATTGTCTAAAGTCCCTGGAAGAGCAGC
TCAAGACTGCTCTGTCTGTAAGTAGAGCTAGGAAGACAGAACTAATGATGGAGCTTGTGA
AGACCCATCAAGAGAAGGAGAAGCTGCTGAGAGAGGAGAACCAGAGTTTGAATAACCAGC
TTATAAAGTAGGGGAAGATGAAGAAGTCTGTGGAAGCAGAGGATGCAAGAGCAATGTCAC
CGGAAAGTAGCTCTGACAAACAGCCACCGGAGACTCTCCTGCTTCTCAAGTAACCACCAT
CACCACGACTGATTGCAAAAAATAAAATTTGTAATAATTATGATTTGTAGTTCATAAGGA
AAGCTACATACTGTATGTTAAAAATCCTTCTTCCCCCTGCTACGGAAGAGTCATCCAA
GGAGATGCATCAAATAAAGTAATTGATTTTATTGTTA

>G1843 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKVEIKRIENKSSRQVTFCKRRNGLMEKARQLSILCESSVALIIISATGRLYSFSSG
DSMAKILSRYLEQADDLKTLDLEKTLNLYLSHKELLETIQCKIEBAKSDNVSIDCLKSL
EEQLKTALSVTRARKTELMELVKTHQEKEKLLREENQSLTNQLIKMGKMKKSVEAEDAR
AMSPSSSDNKPPETLLLLK*

>G1853 (1..1860)

ATGAGAGGTTCTTGGTACAAGAGTGTTTCTCTGTTTTTGGTCTCAGACCACGGATCAGA
GGGTTGTTATTCTTTCATTGTTGGTGTGTGGCTCTAGTTACTATTTTAGCACCATTGACA
TCTAATTCGTATGATTCTTCGTCAAGTTCGACACTTGTGCCGAACATTTATAGTAACTAT
AGGAGGATAAAGGAGCAAGCTGCTGTTGATTATCTTGATCTGAGGTCTCTTTCTTTAGGG
GCTAGTTTTAAAGAGGTTTCTTTTGTGGTAAAGAAAGAGAAAGTTATGTGCCTTGTAT
AACATAACTGGGAATTTGCTTGCTGGGCTTCAAGAGGGTGAGGAGTTAGATCGACATTGC
GAGTTTGAAAGAGAGAAGGAAAGATGTGTAGTTCGTCTCCGAGAGATTATAAAATACCA
CTTAGGTGGCCACTTGGTAGAGATATCATATGGAGTGGGAACGTGAAGATTACCAAAGAC
CAGTTTCTTTCTTCAGGAACGTGACAACGAGGTTAATGTTGCTTGAAGAGAATCAAATA

ACCTTTCACTCGGAGGACGGCCTGGTCTTTGATGGGGTCAAAGACTATGCTCGTCAAAT
GCTGAGATGATAGTTTATAGGAAGTGATACTGAATTTGCTCAAGCGGGTGACGGACTGTG
TTAGACATTGGTTGCGGATTTGGTAGCTTTGGTGCTCATTTAGTGTCTTTGAAGCTGATG
CCTATATGTATTGCTGAGTATGAGGCAACTGGGAGCCAAGTTCAGTTAGCTCTAGAGAGA
GGCCTTCCTGCAATGATTGGCAATTTCTTTTCAAACAGCTTCCTTATCCAGCACTGTCT
TTTGATATGGTCCATTGTGCTCAATGTGGCACTACTTGGGATATCAAAGATGCAATGCTA
CTTTTGGAAAGTGGATCGTGTCTGAAACCCGGGGGATACTTTGTTTAACTTCTCCACA
AACAAAGCACAGGGAACTTACCGATACCAAGAAAACGAGCATCTCAACACGGGTGAAT
GAGTTATCTAAGAAAATCTGTTGGAGTCTAACAGCTCAGCAGGATGAGACGTTTCTTTGG
CAGAAAACCTTCTGATTCAAGTTGCTATTCTTCTCGTTCGCAAGCTTCTATACCTCTTTGC
AAAGATGGAGATAGCGTTCCGTATTACCAACCATTTGGTTCATGTATAAGCGGAACACG
AGTAAACCGCTGGATTTCTATACAGAACAGGTCTGCTGTTGCAGGAACAACCTCTGCCGGG
CTTGAAATTCATGGTTTAAACCCGGAAGAATTCTTCGAGGATACACAAATATGGAGATCA
GCTCTGAAAACTATTGGTCTTGTCTTACACCTCTAATTTTCTCTGACCATCCGAAGAGA
CCCGGTGATGAGGATCCTCTCCCGCCTTTCAACATGATACGCAATGTGATGGACATGCA
GCTCGTTTGGGAAATTTAAATGCCGCTTTACTCGACGAAGGAAAACTGCTTGGGTAAATG
AACGTCGTCCTCAGTCAATGCACGTAATACTCTTCTATCATACTTGATCGTGGTTTCGCC
GGTGTCTACATGACTGGTGTGAACCATTCCTCGACATATCCTCGAACATATGACATGCTT
CATGCCAATGAATTTCTCACACATCTTAGCTCAGAACGATGCAGCCTAATGGACTTGTTCT
TTGGAGATGGACCGGATTCTTCGCCCTGAGGGATGGGTGTTCTAAGCGACAAAGTGGGA
GTAATCGAGATGGCTCGACACTTGCAGCTCGAGTGCCTTGGGAAGCAAGAGTCATTGAT
CTTCAAGATGGTAGTGACCAAGACTTCTCGTCTGTCAAACCATTCATCAAAAAATAA
>G1853 Amino Acid Sequence (domain in AA coordinates: entire protein)
MRGSWYKSVSSVFLRPRIRGLLFFIVGVVALVTILAPLTSNSYDSSSSSTLVPNIYSNY
RRIKEQAAVDYLDLRLSLGASLKEFPFCGKERESYVPCYNITGNLLAGLQEGEELDRHC
EFEREKERCVVRPRDYKIPLRWPLGRDIIWSGNVKITKDQFLSSGTVTTRLMLLEENQI
TFHSEDLGVFDGVKDYARQIAEMIGLSDTEFAQAGVRTVLDIGCCGFSFGAHLVSLKLM
PICIAEYEATGSQVQLALERGLPAMIGNFFSKQLPYPALSFDMVHCAQCQTWDIKDAML
LLEVDRLVLPKGGYFVLTSPTNKAQGNLPDTHKTSISTRVNELSKKICWSLTAQQDETFW
QKTSDDSSCYSSRSQASIPCKDGDSPYHPLVPCISGTTSKRWISIQNRSAVAGTTSAG
LEIHGLKPREEFFEDTQIWSALKNYWSLLTPLIFSDHPKRPGEDEPLPPFNMRNVMDMH
ARFGNLNAALLDEGKSAWVMNVVFNARNTLPILDRGFAGVLHDWCEPFPTYPRTYDML
HANELLTHLSSERCSLMDLFLEMDRILRPEGWVVLSDKVGVIEMARALAAVRVWEARVID
LQDGSQDRLLVCKPFKK*
>G1855 (1..1902)
ATGGCGAAAGAGAACAGTGGTCATCATCACCAAACAGAAGCAAGAAAGAAAGAACTAACT
TTGATTCTTGGTGTAAAGTGGACTCTGCATTTTGTCTATGTTTATAGGTGCATGGCAAGCC
AATACCGTCCCATTCTATCTCGAAGCTCGGATGCGAGACGCAATCAAACCTTCTTCG
TCCTCTTCTCTTCTCATCTTCAGAGTCAGCTGAAGTAGATTCAAAGCCATAATCAG
ATTGAGTTAAAGGAAACAAACCAACCATTAAGTACTTTGAACCATGTGAATTATCTCTC
AGTGAGTACACTCCTTGTGAAGACCGACAAAGAGGAAGAAGATTGATAGGAACATGATG
AAATATAGAGAAAGACATTGTCTGTAAAAGATGAGCTTCTTATTGTTTGATTCTCCT
CCACCAAACACAGATTCCATTTAAATGGCCACAAAGTAGAGACTATGCTTGGTATGAC
AATATCCCTCACAAAGAACTTAGTGTGAGAAAGCAGTTCAAACTGGATTCAAGTTGAA
GGTGACCGCTTTAGATTCCCTGGTGGTGGTACTATGTTTCTCGTGGAGCTGATGCTTAT
ATCGATGATATTGCTAGGCTTATCTCTTACTGATGGTGAATCAGAACAGCTATTGAC
ACTGGATGTGGTGTGCAAGTTTGGTGCTTACCTCTTGAAGAGAGACATTATGGCTGTG
TCTTTTGTCTCCAAGAGACTCATGAAGCTCAGGTACAGTTTGCTTTAGAACCGCGAGTT
CCTCGGATAATCGGATATGGGATCAAGAAGACTTCTTATCCAGCTAGAGCTTTTGAT
CTTGCTCATTGTTCTCGTTGTGTTGATCCCTTGGTTTAAAAATGATGGTTTGTACCTTATG
GAGGTCGACCGGTTTAAAGACCGGGCGGTTACTGGATCCTCTCGGGACCACCGATTAAAC
TGGAAACAGTACTGGAGAGGGTGGGAGAGAACAGAGGAGGATTTGAAGAAAGAGCAAGAT
TCAATAGAAGATGTAGCAAAGAGTCTTTGCTGGAAGAAAGTAACTGAAAAAGGTGACTTA
TCAATTTGGCAAAAGCCTCTCAATCACATTGAGTGTAAGAGCTCAAAACAAAACAATAAG
TCACCTCCGATATGCAGTACAGATAACGCGGATTCCGCTTGGTACAAAGACTTGGAACT
TGATAACACCAATTACCGAGAAACAAACAATCCAGATGATTACAGAGCGGTGCACTCGAG
GATTGGCCAGACCGAGCATTCGCGGTACCTCCAAGAATCATCAGAGGAACATATACCAGAA

ATGAACGCGGAGAAATTTAGAGAAGACAACGAGGTTTGGAAAGAGAGAATAGCACATTAC
AAGAAGATAGTCCCTGAGCTTTACATGGAAGATTACGGAACATTATGGACATGAACGCT
TTTCTCGGCGGATTTCGCTGCTTCCATGCTGAAATATCCCTCATGGGTCATGAACGTTGTC
CCGGTCGATGCAGAGAAACAAACGTTAGGTGTGATCTACGAACGTGGATTGATAGGGACG
TATCAAGATTGGTGTGAAGGATTCTCAACGTATCCAAGAACTTATGATATGATTTCATGCA
GGAGGATTGTTTCAGCTTATACGAACATAGGTGTGATTTGACGTTGATATTGTTGGAGATG
GATCGAATTTTTCAGACCAGAAGGAACAGTTGTGTTGAGAGATAATGTGGAGACGTTGAAT
AAGGTAGAGAAGATAGTGAAGGGAATGAAGTGAAGAGTCAAATTGTTGATCATGAGAAA
GGTCCTTTTAACTCCTGAGAAGATTCTTGTGCTGTTAAACTTATTGGACTGGTCAACCT
TCTGACAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA

>G1855 Amino Acid Sequence (domain in AA coordinates: entire protein)

MAKENSQHQQTEARRKKLTLLILGVSLCILFYVLGAWQANTVPSSISKLGCTQSNPSS
SSSSSSSESSELDFKSHNQIELKETNQTIKYFEPCELSLSEYTPCEDRQGRFRFDRNMM
KYRERHCPVKDELLYCLIPPPNYKIPFKWPQSRDYAWYDNIPHKELSVKAVQNWIQVE
GDRFRFPGGGTMFPRGADAYIDDIARLIPLTDGGIRTAIDTGCGVASFGAYLLKRDIMAV
SFAPRDTHEAQVQFALERGVPAIIGIMGSRRLPYPARAFDLAHCRCCLIPWFKNDGLYLM
EVDRLVRPGGYWILSGPPINWKQYWRGWERTBEDLKEQDSIEDVAKSLCWKKVTEKGD
SIWQPLNHIIECKLKLQNNKSPICSSDNADSAWYKDLCTITPLPETNNPDDSGAGLE
DWPDRFAFVPPRIIRGTIPEMNAEKFPREDNEVWKERIAHYKKIVPELSHGRFRNIMDMNA
FLGGFAASMLKYPWSVMNVVPDAEKQTLGVIYERGLIGTYQDWCEGFSTYPRTYDMIHA
GGLFSLYEHRCDLTLILLEMDRILRPEGTVVLRDENVETLNKVEKIVKGMKWSQIVDHEK
GPFNPEKILVAVKTYWTGQPSDKNNNNNNNNNN*

>G187 (118..1074)

TAGACCTCTTAGGAAAAAACCTAAAAACCTAATCCCCAACCTAAAAGGCTTATCTCAT
CTCTTCTCTTTGTTCTTTACTCTTTTTTTTACCTCTCTCTCATTGTTCTTACCATG
TCTAATGAAACCAGAGATCTCTACAACCTACCAATACCCTTCATCGTTTTCGTGCACGAA
ATGATGAATCTGCCTACTTCAAATCCATCTTCTTATGGAACCTCCCATCACAAAACGGT
TTTAATCCATCTACTTATTCCTTCACCGATTGTCTCAAAGTTCTCCAGCAGCGTATGAA
TCTCTACTTCAGAAAACCTTTTGGTCTTTCTCCCTCTTCTCAGAGGTTTTCAATTCTTCG
ATCGATCAAGAACCAGAACCGTGATGTTACTAATGACGTAATCAATGGTGGTGCATGCAAC
GAGACTGAAACCTAGGTTTCTCCTTCTAATTCTTCTCTAGTGAGGCTGATCACCCCGGT
GAAGATTCCGGTAAGAGCCGGAGGAAACGAGAGTTAGTCCGGTGAAGAAGATCAAATTTCC
AAAAAAGTTGGGAAAACGAAAAGACTGAGGTGAAGAAACAAAGAGAGCCACGAGTCTCG
TTTATGACTAAAAGTGAAGTTGATCATCTTGAAGATGTTATAGATGGAGAAAATACGGC
CAAAAGGCTGTAAAAATAGCCCTTATCCAAGGAGTTACTATAGATGTACAACACAAAAG
TGCAACGTGAAGAAACGAGTGGAGAGATCGTTCCAAGATCCAACGGTTGTGATTACAAC
TACGAGGGTCAACACAAACCCGATTCCGACTAATCTTCGAGGAAGTTCTGCCGCGGCT
GCTATGTTCTCCGACACCTCATGACTCCAAGAAGCTTGCACATGATATGTTTAGGACG
GCAGCTTATACTAACGGCGGTTCTGTGCGCGGCTTGGATTATGGATATGGACAAAGT
GGTTATGGTAGTGTGAATTCAAACCCTAGTTCTCACCAAGTGTATCATCAAGGGGGTGAG
TATGAGCTCTTGAGGGAGATTTTCTCTCAATTTCTTAAAGCAAGAGCCTTGATCGATC
ATTGTTATAACTACATATATTATATATATATGAGAGAGAGAGGTAGAGAAAAA

>G187 Amino Acid Sequence (domain in AA coordinates: 172-228)

MSNETRDLYNQYPSSFSLEHMMNLPTSNPSSYGNLPSQNGFNPSTYSFTDCLQSSPAAY
ESLLQKTFGLSPSSSEVFNSSIDQEPNRDVTNDVINGGACNETETRVSPSNSSSEADHP
GEDSGKSRRLRELVEEDQISKVVGKTKKTEVKKQREPRVSFMTKSEVDHLEDGYRWRKY
GQKAVKNSPYPRSYRCCTTQKCNVKKRVERSFQDPTVVITYEGQHNHPIPTNLRGSSAA
AAMFSADLMTPRSFAHDMFRTAAYTNGGSVAAALDYGYSQSGYGSVNSNPSSHQVYHQGG
EYELLREIFPSIFFKQEP*

>G1881 (1..519)

ATGCGAATTTTGTGTGATGCTTGTGAGAGCGCCGCCGCTATCGTCTTTTGGCGCCCGAC
GAAGCTGCCCTCTGTTGCTCTGCGACGAAAAAGTTCAAGTGCAACAAGCTGGCTAGT
CGGCATCTTCGTGTAGGCTTAGCTGATCCGAGTAATGCACCAAGCTGTGACATATGCGAA
AATGCACCCGATCTTTTACTGTGAGATAGATGGTAGTTCCCTTTGTCTACAATGTGAT
ATGGTGGTACATGTTGGTGGGAAGAGAACACATAGGCGGTTTCTATTACTGAGACAGAGA
ATTGAGTTTCCAGGCGATAAGCCTAATCATGCTGACCAACTGGGACTACGGTGTCAAAG
GCTTCTCTGCTGCTGGTCAAGAATCAAATGGGAATGGTGATCATGATCATAATATGATC

GATCTTAACTCCAATCCTCAAAGAGTACACGAGCCTGGATCACATAACCAAGAGGAGGGT
ATTGATGTAAATAACGCAAACAATCAGGAGCATGAATAG
>G1881 Amino Acid Sequence (domain in AA coordinates:5-28, 56-79)
MRILCDACESAAAIVFCAADEAALCCSCDEKVHKCNKLASRHLRVGLADPSNAPSCDICE
NAPAFYCEIDGSSLCLQCDMVVHVGGKTRHRRFLLLRQRIEFPDGPKNHADQLGLRCQK
ASSGRGQESNGNDHDHNMIDLNSNPQRVHEPGSHNQEEGIDVNNANNHEHE*
>G1882 (1..1200)
ATGGTTTTTCTTCATTTCTTACTTATCCTGATCATTTCATCAAACCTGGCAACAACAACAT
CAACCAATCACAAACCACCGTTGGATTACAGGGAAATAACATCAACCAACAGTTTCTTCCT
CACCATCCCCCTCCACCGCAACAGCAACAAACGCCTCCGCAGCTTCACCACAACAACGGT
AACGGCGGAGTCGTGTTCCCGTGGACCTGGCGGGTTAATCCGACCAGGTTTCGATGGCG
GAAAGAGCAAGGCTAGCCAACATACCATTACCTGAAACAGCCTTGAAGTGTCCAAGATGT
GACTCAACTAACACCAAATTCGTGTTACTTCAACAACTACAGTCTCACTCAACCTCGCCAC
TTCTGCAAAGCATGCCGTGTTACTGGACACGTGGCGGTGCTCTAAGGAGCGTTCCCGTC
GGTGGCGGTTGCCGTAGAAACAAAAGAACAAAAACAGCAGCGGTGGAGGTGGCGGTAGC
ACCAGTAGCGGTAAACAGCAAGTCACAAGACAGCGCCACGAGCAACGACCAATACCACCAC
CGAGCCATGGCTAACAAATCAGATGGGACCACCTTCTTCGTTCATCGTCTCTAAGCTCGTTG
CTGTCTTCTTACAACGCAGGGTTAATCCCCGACATGATCATAACAGCAATAACAACAAC
ATACTTGGACTTGGATCATCTTTGCCTCCTCTTAAGCTTATGCCTCCTTTAGACTTCACA
GACAACTTCACCTTACAATACCGTGCCGTTTCAGCTCCTTCTTATCATATAGCGCGTGGA
AGCAGTGGAGGAGCGCGGCTCTTTTAAACGGTTTGGACAGTGGAGATTCCCGGCAACA
AACCAACTTCTTTAGGCGGTTTAGACCCGTTTGATCAACAACATCAAATGGAGCAGCAG
AATCCAGGTTACGGATTGGTTACCGGGTCGGGTTCAGTATCGACCTAAGAACATTTTCCAT
AACCTTATCTCCTTCTTCTCGTCTGCTTCATCAGCTATGGTTACAGCCACCGCGTCGCAA
TTAGCTTCAGTGAAATGGAAGATAGTAACAATCAGCTCAACTTGTCTAGACAACCTTTT
GGAGACGAACAACAGCTCTGGAATATTTCATGGCGCTGCTGCAGCATCCACCGCAGCTGCA
ACAAGTTCGTGGAGTGAAGTCTCTAATAATTTTCAGTTCTTCTTCTACTAGCAATATATAA
>G1882 Amino Acid Sequence (domain in AA coordinates:97-125)
MVFSFPTYPDHSSNWQQHQPIITTVGFTGNINQQFLPHHPLPPQQQQTTPQLHHNNG
NGGVAVPGGPGGLIRPGSMAERARLANIPLPETALKCPRCDSTNTKFCYFNYSLTQPRH
FCKACRRYWTRGGALRSVPVGGGCRNRKTKNSSGGGGSTSSGNSKSDSATSNDQYHH
RAMANNQMGPSSSSSSLLSSYNAGLIPGHDHNSNNNNILGLGSSLPPLKMPPLDFT
DNFTLQYGAVSAPSYHIGGGSSGAAALLNGFDQWRFPATNQLPLGGLDPFDQHQHMEQQ
NPGYGLVTGSGQYRPKNIFHNLISSSSSASSAMVTATASQLASVKMEDSNNQLNLSRQLF
GDEQQLWNIHGAAAAATAATSSWSEVSNFSSSSTSNI*
>G1883 (1..1110)
ATGGACGCTACGAAGTGGACACAGGGTTTTCAAGAAATGATGAACGTTAAACCAATGGAG
CAGATCATGATTCCTAATAACAACACACATCAACCAACACCACATCCAATGCAAGGCCA
AACACCATTTCTCATCTAACGGCGTCTCAACTGCTGGAGCAACCGTCTCCGGCGTAAGC
AACACAATAACAATACGGCGGTTGTGGCGGAGAGGAAAGCAAGACCACAAGAGAACTA
AATGTGCAAGATGCAACTCAACCAACACAAAGTTTTGTTACTACAACAATATAGTCTC
ACACAACCAAGATACTTCTGCAAAGGTTGTGGAAGGTATTGGACCGAAGGTGGATCTCTT
AGGAATGTTCTGTGGGAGGAAGCTCAAGAAAGAACAGAGATCATCTTCATCTTCTTCA
TCAAACATCCTTCAGACAATACCATCTTCACTTCCAGATCTAAACCCGCCAATACTCTTC
TCAAACCAATCCATAATAAATCGAAAGGGTCATCACAAGATCTCAACTTGTGTCTTTC
CCAGTCATGCAAGATCAACATCATCATCATGTCCATATGTCTCAGTTTCTTCAGATGCCT
AAGATGGAGGGAAATGGTAACATAACTCATCAGCAGCAGCCTTCATCATCTTCTTCTGTC
TATGGTTCCTCGTCCTCTCTGTTTCAGCTCTTGAACTTTTAAGAACCGGAGTTAATGTT
TCTTCAAGATCAGGGATTAACATCATCGTTCATGCCTTCCGGTTCAATGATGGATTCAAAC
ACTGTGCTTTTACCTTCTTTCAGGGTTTCCAACAATGGTGGATTACAAGCCAAGTAATCTC
TCCTTCTCTACCGATCATCAAGGGCTTGGACACAATAGCAACAATAGGTCTGAAGCTCTT
CATAGTGATCATCACCAACAAGGTAGAGTTTTGTTTCCATTGGGGATCAAATGAAGGAG
CTTTCATCAAGCATAACACAAGAAGTTGATCATGATGATAATCAACAACAGAAGAGTCAT
GGAAATAATAATAATAATAATAACTCAAGCCCTAATAATGGATATTGGAGTGGGATGTTT
AGTACTACAGGAGGAGGATCTTCATGGTGA
>G1883 Amino Acid Sequence (domain in aa coordinates: 82-124)
MDATKWTQGFQEMNVKPMEQIMIPNNNTHQPNTTSNARPNTILTSNGVSTAGATVSGVS

NNNNNTAVVAERKARPQEKLNCPRCNSTNTKFCYNNYSLTQPRYFCKGCRRYWTEGGSL
RNVPVGGSSSRKNKRSSSSSSSSNILQTI PSSLPDLNPPILFSNQIHNKSKGSSQDLNLLSF
PVMQDQHVVHMSQFLQMPKMEGNGNITHQQQPSSSSSVYSSSSPVSALELLRTGVNV
SSRSGINSSFMPSGSMDSNTVLYTSSGFPTMVDYKPSNLSFSTDHQGLGHNSNNRSEAL
HSDHHQQGRVLFPPFDQMKELSSSITQEVDDHNNQQQKSHGNNNNNNSSPNNGYWSGMF
STTGGGSSW*

>G1884 (1..741)

ATGATGACGTCATCCCATCAGAGCAACACCACCGGCTTTAAACCGCGGCGGATCAAGACG
ACGGCGAAGCCACCACGTCAGATCAATAACAAAGAACCATCTCCGGCGACGCAGCCGGTG
CTCAAGTGTCAGAGATGTGATTCAAGTCAACACCAAATCTGCTACTACAACAACACTACAGC
TTGTCTCAGCCACGTCCTACTGCAAGAACTGTCTGCTTACTGGACACGTGGCGCGGCC
CTCCGTAACGTTCCCATCGGTGGCTCCACTCGAAACAAGAACAAGCCTTGCAGCCTCCAA
GTCATCTCTTCTCTCTTTGTCTCGAACGGGACGTCATCGGCGTCTCGTGAGCTTGTA
AGAAACCATCCATCGACGGCAATGATGATGATGAGTTCTGGTGGATTCTCCGGCTATATG
TTTCCGTGGATCCTAACTTCAACCTTGCCTCGTCTTCTATCGAGTCTTTGAGTTCTTTT
AACCAAGATTGCAACAGAAGCTTCAGCAACAAAGACTCGTCACTTCCATGTTTCTCCAA
GATTCTCTTCCGGTTAACGAGAAAACGGTTATGTTTCAGAACGTAGAGTTGATTCTCTCT
TCGACGGTGACGACGATTGGGTTTCGATAGGTTGCCACTGGAGGAGGTGCAACAAGT
GGCAATCATGAAGATAATGATGATGATGGGGAGGGTAATTGGGAAATGGTTCCATAATGCT
AATAATAATGCTCTGCTCTAA

>G1884 Amino Acid Sequence (domain in AA coordinates:43-71)

MMTSSHQNTTGFKPRRIKTTAKPPRQINNKEPSPATQPVLKCPRCDSVNTKFCYNNYS
LSQPRHYCKNCRRYWTRGGALRNVPIGGSTRNKNKPCSLQVISSPPLFSNGTSSASREL
RNHPSTAMMMSSGGFGSYMFPLDPNFNLAASSIESLSSFNQDLHQKLOQORLVTSMFLQ
DSLVPNEKTVMFQNVELIPPSTVTTDWVFDRAFATGGGATSGNHEDNDDEGNLGNWFHNA
NNNALL*

>G1891 (1..750)

ATGGATAACTTGAATGTTTTTCGCAAATGAAGACAATCAAGTGAATGATGTGAAGCCCCA
CCACCACCACCTCGAGTGTGTGCAAGGTGTGATTCTGATAATACTAAATTTTGTTATTAC
AACAACTACTGTGAGTTTCAGCCACGATACTTCTGCAAGAACTGTCGTAGATACTGGACT
CATGGTGGGGCTTTAAGAAACATACCAATTGGTGGAGTAGTCGTGCCAAACGGGCAAGG
GTAAATCAACCTTCGGTTGCTCGGATGGTTTCTGTTGAGACCAACGAGGTAACAATCAA
CCTTTCTCTAATGTTCAAGAAACGTTTCATCTTGTGGATCTTTTGGTGCTTCATCTTCA
TCTTCTGTGGTGCTGTGGGAACCTTTTGGTCTTTGTATGATATTCATGGTGGTATG
GTAACAAATTTGCATCCAACCTCGAATGTTTCGACCAATCATCGCTTAGCTTTCCATGAT
GGATCATTTGAGCAAGACTATTACGATGTTGGGTCCGATAATCTTTGGTCAACCAACAA
GTTGGTGGCTACGGTTATCACATGAATCCAGTGGATCAATTCAAGTGGAAACCAGAGCTTC
AACAACTATGAACATGAATTATAATAACGATAGCACTAGTGAAGTAGCAGAGGATCT
GACATGAATGTGAACCATGATAACAAGAAGATCAGATACCGCAACTCTGTGATTATGCAT
CCTTGTCTCTGGAGAAGGATGGTCTTGA

>G1891 Amino Acid Sequence (domain in aa coordinates: 27-69)

MDNLNVFANEDNQVNDVKPPPPPRVCARCDSDNTKFCYNNYCEFPQRYFCKNCRRYWT
HGGALRNIPIGGSSRAKRARVNQPSVARMVSVETQRGNNQPFQSNVQENVHLVGSFGASS
SSVGAVGNLFGSLYDIHGGMVTNLHPTRTRVPRNHLAFHDGSFEQDYDVGSNDLLVNQQ
VGGYGYHMPVDQFKWNQSFNNTMNMNYNDSTSGSSRSGSDMNVNHDNKKIRYRNSVIMH
PCHLEKDG*

>G1896 (1..951)

ATGTCCTCCCATAGCAATCTCCCCTCTCCCAAACAGTTCCTAAACCAGATCACCGTATC
TCCGGTACATCCCAAACCAAGAAACCACCGTCTTCTCCGTAGCTCAAGACCAACAAAAC
CTAAATGCCCTCGTTGCAACTCTCCAAACACAAAGTTCTGTTACTACAACAACACTACAGT
CTCTCTCAACCTCGTCACTTCTGCAATCTTGTGCGCGTTACTGGACACGTGGCGGTGCT
CTAAGAAACGTCCCCATCGGTGGTGGTTGCCGGAACCAAAAATCTATCAAACCTAAT
TCCTCCATGAACAACTTCTTCTTCTTCTTCTCTCAGAGGTTCTTCTCATCAATCATG
GAAGATTCATCCAAATCTTCCCTCCTCCGACAAATGATTTTTCAGCTGGCCGGATTA
TCTCTCAACAAAATGAACGATCTTCAACTTTTGAATAACCAAGAAGTTCTTGATCTTAGG
CCCATGATGTCCTCGGGCCGAGAAAACACACCCGTTGATGTCGGGTGGGTTTATCCCTA
ATGGGTTTTGGAGATTTCAACAACAACCATCACCGACGGGGTTCACAACCGCCGGAGCA

AGCGACGGAAACTTAGCTTCTTCTATAGAGACTTTGAGTTGTTTAAACCAAGATTTACAC
TGGAGGCTTCAGCAACAGAGGATGGCGATGCTTTTTGGTAATTCCTAAGGAAGAACTGTT
GTCGTCGAGAGGCCACAACCTATTCTTTATCGGAATCTTGAGATCGTAAACTCATCATCG
CCGTCGTCGCCGACGAAGAAAGGAGATAATCAGACAGAGTGGTATTTTGGTAATAACAGT
GATAATGAAGGAGTGATTAGTAATAATGCTAATACAGGAGGAGGAGGAAGTGAATGGAAC
AATGGAATTCAAGCTTGGACTGATCTTAATCATTATAATGCATTGCCTTGA

>G1896 Amino Acid Sequence (domain in aa coordinates: 43-85)
MSSHTNLPSPKPVKPDHRISGTSQTKKPPSSSSVAQDQQLKPCRCNSPNTKFCYNNYS
LSQPRHFCKSCRRYWRGGLRNVP.IGGGCRKTKKSIKPNSSMNTLPSSSSSQRFSSIM
EDSSKFFPPPTTMDFLAGLSLNKMNDLQLLNNQEVLDLRPMSSGRENTPVVDVGSGLSL
MGFGDFNNHNSPTGFTTAGASDGNLASSIETLSCLNQLDHLWRLQQORMAMLFNGSKEETV
VVERPQPILYRNLEIVNSSSPSSPTKKGDNQTEWYFGNNSDNEGVISNNANTGGGSEWN
NGIQAWTDLNHYNALP*

>G1898 (1..630)
ATGCCGTCGGAACCAACCAACCCGACCCACCAGAGTTCAGCCCTCAACGGCGGCTTAC
CCACCGCCAAATCTGGCTGAGCCTCTTCCTTGTCCTCGCTGCAACTCCACCACCACCAAG
TTCTGTACTACAACAATAACCTCGCTCAGCCTCGCTACTACTGCAAATCTTGCCGC
CGTTACTGGACTCAAGTGGTACACTCCGTGACGTCCCGTCGGTGGTGGAACTCGTCGA
AGCTCCTCAAAACGTCACCGTTCTTTCTCCACCCTGCCACCTCTCTCTCTCTCTCT
TCCGTCAATCACCACCACGACACAAGAACCAGCCAGACTGAAGCGAGTCAAACCTAAGGTT
ACTAATTTAATTTAGTTCATGGAAGCTTTGCTTCTCTGTTAGGTTTAGGAAGTGGAAT
GGTGGGTTGGATTACGGGTTTGGGTACGGGTACGGGCTTGAGGAGATGAGTATTGGGTAT
CTTGGAGATTCTTCCGTAGGAGAGATTCCGGTGGTTGATGTTTGGTGGTGACACGTGG
CAGATTGGGGAGATTGAAGGTAAAGTGGAGGAGACAGTTTGATATGGCCTGGTCTTGAG
ATCTCAATGCAAACCAACGATGTTAAGTGA

>G1898 Amino Acid Sequence (domain in AA coordinates: 31-59)
MPSEPNTQTRPVRQSTAAAYPPPNLAELPCPRCNSTTKFCYNNYNLAQPRYCKSCR
RYWTQGGTLRDVVPVGGGTRRSSSKRHSFSTTATSSSSSSSVITTTTQEPATTEASQTKV
TNLISGHGSFASLLGLGSGNGGLDYGFYGYGLEEMSIGYLGDSVGEIPVVDGCGGDTW
QIGEIEGKSGGDSLIWPGLEISMQTNDVK*

>G1902 (1..615)
ATGCAGGATCCAGCAGCATATTACCAGACGATGATGGCGAAGCAACAACAACAACA
CCACAGTTTGCAGAGCAAGAACAGTTAAAGTGTCTCGTTGTGACTCACCAAACACTAAA
TTCTGTACTACAACAATACTCTCACAGCCTCGTCACCTTTTGCAAAAGCTGTCTGT
CGTTACTGGACTAAAGCGCGCTCTCCGTAACGTTCCCGTCGGTGGTGGTCTCGTAAG
AACGCAACCAACGATCCACTTCTTCTTCTTCTGCTTCTCTCTTCCAACAGTAGC
CAAAACAAGAAGACGAAAAACCCGGATCCGGATCCTGATCCAGTAATTCTCAAAAACCG
GATTTGGATCCGACCCGGATGCTTTACGGGTTTCCGATCGGTGACCAAGACGTGAAGGGT
ATGGAGATTGGTGGAGCTTTAGCTCGTTGTTGGCGAATAATATGCAGCTTGGTCTTGGA
GGAGGAGGGATCATGCTTGACGGGTCCGGTTGGGATCATCCGGTATGGGTTTGGGTTTG
AGGAGAACCGAACCGGTAATAATAATAATAAACCATGGACCGATCTGGCTATGAACAGA
GCGGAGAAAAAAGTGA

>G1902 Amino Acid Sequence (domain in AA coordinates:31-59)
MQDPAAYYQTMMAKQQQQQPFQFAEQQLKPCRCDSPTKFCYNNYNLSQPRHFCKSCR
RYWTKGGLRNVPVGGGSRKNATKRSTSSSSSASSPSNNSQNKKTKNPDPPDPNRSQKP
DLDPTRMLYGFPIGDQDVKGMEIGGSFSSLLANNMQLGLGGGIMLDGSGWDHPPMGLGL
RRTEPGNNNNNPWTDLAMNRAEKN*

>G1904 (1..924)
ATGCAAGATATTCTGATTTCTCCATGAACGGAGTTGGTGGTGGGGGAGGAGGAGGGG
AGGTTTTTTCGGTGGAGGAATCGGCGGCGGAGGAGGTGGTGATCGAAGGATGAGAGCTCAT
CAGAACAATATACTTAACCATCATCAATCTCTCAAGTGTCTCGTTGTAATTTCTTTAAC
ACAAAGTTCTGTACTACAACAATTACAATCTTTCTCAGCCTCGTCACCTTTTGCAAGAAC
TGTCGTCGTTACTGGACTAAAGGTGGTGTCTCCGTAACGTTCCCGTCGGAGGTGGTTGC
CGGAAAGCTAAACGTTTGAAACAAAACAGGTTCCGTCGTCGTCATCAGCCGACAAACCA
ACGACGACGCAAGATGATCATCAGTGGAGGAGAAATCGAGTACAGGATCTCACTTAGC
AGCGAGAGCTCTTCTCTCACCGCTTCTAACTCTACCACCGTCGCCGCCGCTCTCCGTACCC
GCGGCGGCGGAAGTTGCTTCGTCGGTTATTCCAGGTTTTGATATGCCTAATATGAAAATT

TACGGTAACGGGATCGAGTGGTTCGACGTTACTTGGACAAGGCTCATCGGCCGGTGGTGTT
TTCTCGGAGATCGGTGGTTTTCCGGCGGTTTCAGCTATTGAACTACACCGTTTGGATT
GGGGGTAAATTCGTAAATCAAGATGATCATCTGAAGTTAGAAGGTGAACTGTACAGCAG
CAACAGTTTGGAGATCGAACGGCTCAGGTTGAGTTTCAAGGAAGATCTTCGGATCCGAAT
ATGGGATTTGAACCGTTGGATTGGGGAAGTGGCGGTGGAGATCAAACACTGTTTGATT
ACCAGTACCGTTGATCATGCTACTGGAGTCAAAGTCAATGGACGTCGTCTGACCAAGAT
CAGAGTGGTCTCTACCTTCCTGA

>G1904 Amino Acid Sequence (domain in aa coordinates: 53-95)
MQDIHDFSMNGVGGGGGGGRFFGGGIGGGGGDRRMRAHQNNILNHHQSLKPCRCNSLN
TKFCYYNNYNLSQPRHFCKNCRRYWTGGVLRNVPVGGGCRKAKRSKTKQVPSSSSADKP
TTTQDDHHVEEKSSTGSHSSSESSSLTASNSTTVAAVSVTAAAEVASSVIPGFDMPNMKI
YNGIEWSTLLGQSSAGGVFSEIGGFPAVSAIETTPFGFGGKFVNQDDHLKLEGETVQQ
QQFGDRTAQVEFQGRSSDPNMGFEPLDWGSGGGDQTLFDLTSTVDHAYWSQSQWTSSDQD
QSGLYLP*

>G1906 (1..795)
ATGGTGGAACGTGCTCGGATCGCAAAAGTCCCATTGCCTGAAGCAGCTCTAAATTGCCCT
AGATGTGACTCAACCAATACTAAGTTCTGTACTTCAATAACTATAGCCTTACTCAACCT
CGCCATTTCTGCAAAACATGTCGTGCTATTGGACACGTGGCGGTTCTTGAGGAATGTT
CCTGTTGGAGGAGGCTTTAGGAGGAACAAGAGAAGCAAATCCAGATCGAAATCTACGGTC
GTGGTCTCGACTGATAATACTACTAGTACTTCATCACTTACTTCTCGCCCAAGTTACTCA
AACCTTAGCAAGTTTCATAGCTACGGTCAAATCCCGGAGTTTAAATCCAACTTGCCCATC
TTGCCTCCTCTCCAAAGCCTTGGAGATTACAATTCAAGCAACACTGGATTAGATTTTGGT
GGAACCTCAAATAAGCAACATGATAAGTGGTATGAGTTCTAGTGGTGGGATCTTGGATGCA
TGGAGAATACCTCCATCACCAACAAGCTCAGCAATTCCCTTCTTGATCAACACTACCGGA
TTGGTGCAATCTTCAAACGCGTTATATCCATTACTAGAAGCGGGGTTAGCGCCACGCAA
ACAAGAAATGTGAAGGCGGAAGAGAATGATCAGGATCGGGGTAGGGATGGGGATGGAGTG
AATAACTTATCAAGAACTTTTGGGTAATATCAACATAAATCAGGCAGGAACGAGGAA
TACACATCATGGGAGGTAACAGTTCTTGGACCGGTTTCACTCCAACAACTCAACAGGC
CATCTCTCATTCTAA

>G1906 Amino Acid Sequence (domain in AA coordinates:19-47)
MVERARIAKVPLPEAALNCPDSTNTKFCYFNYSLTQPRHFCKTCRRYWTRGGSRLNV
PVGGGFRNRKRSKRSKSTVVVSTDNSTSSSLTSRPSYNSPKFHSYQIPEFNSNLPI
LPPLQSLGDYNSNTGLDFGGTQISNMISGMSSSGILDWRIPPSQQAQQFPFLINTTG
LVQSSNALYPLLEGVSATQTRNVKAEENDQDRGRDGDGVNLSRNLGNINININSRNEE
YTSWGNSSWTGFTSNSTGHLSF*

>G1913 (1..744)
ATGGAGAGAGCAGGCGCTTGACATCATCGTTTATATGGCGGCCAAACGCAACGCAAC
GCGGAGATCAGCGCGAGTTGTCCAAGATGTGGATCCTCTAACACAAAGTTCTGTACTAC
AACAACTATAGCCTCACTCAGCCTCGCTACTTCTGCAAAGGCTGCCGCAGATATTGGACC
AAAGGTGGTTCCTCCGCAATGTTCTGTAGGCGGTGGCTGTGAAAATCCCGCCGCCCC
AAATCATCTTCTGGTAACAATACTAAACTAGCCTAACCGCTAATCTGGCAACCCCGGT
GGTGGTTCACCAAGCATCGATCTTGCTCTTGTTTACGCCAATTTCTTGAATCCAAAGCCT
GACGAATCTATACTACAAGAAAATGCGACTTAGCCACTACGGATTTTTTGGTAGATAAT
CCTACCGGCACCTCCATGGACCTTCATGGAGTATGGACATCAATGATGGTCATCATGAT
CATTATATTAATCCGGTGGAACACATTGTGGAGGAATGTGGTTATAATGGCTTGCTCCA
TTTCTGGTGAAGAGCTTCTCTTTAGACACTAATGGTGTGTTGGTCTGATGCTTTGTTG
ATTGGTCATAACCATGTAGACGTTGGCGTGACTCCGGTTCAGGCTGTACACGAACCGGTG
GTTCAATTCGCTGAAGAATCCAATGATTCACCAATCTCTGTTTGAAGTTGGAGCCCT
TTTGATTTCACTGCCGATGGATGA

>G1913 Amino Acid Sequence (domain in AA coordinates: 27-55)
MERAELTSSFIWRPNANANAEITPSCPRCGSSNTKFCYYNNYSLTQPRYFCKGCRRYWT
KGGSLRNVPGGGCRKSRPKSSGNNTKTSLTANSNPGGGSIDLALVYANFLNPKP
DESILQENCDLATDFLVDNPTGTSMDPSWSMDINDGHHDHYINPVEHIVEECGYNGLPP
FPGEELLSLDTNGVWSDALLIGHNHVDVGVTVPQAVHEPVVHFADESNDSTNLLFGSWSP
FDFTADG*

>G1914 (1..945)
ATGGAGAGATACAAGTGTAGATTTTGCTTCAAGAGCTTCATCAATGGAAGAGCTTTAGGT

GGTCACATGAGATCTCACATGCTTACTCTTTCTGCAGAACGTTGTGTAATAACTGGTGAA
GCAGAAGAAGAAGTAGAGGAACGGCCGAGTCAACTCTGTGACGACGACGATACCGAG
TCCGATGCTTCTTCTTCTTCTGGTGAGTTTGATAATCAAAAGATGAATCGTCTTGATGAT
GAATTGGAGTTTGATTTCGCTGAAGACGACGACGTTGAAAGTGAAACCGAGTCGTCCAGG
ATTAACCCAACTCGGCGACGATCTAAACGAACTCGGAAACTTGATCGTTTGATTTCGAC
TTTGAGAAGCTAACCAACGAGCCAACCCAGTGAGTTAGTGGCCGAGCCAGAGCATCACAGC
TCAGCTTCTGATACAACAACGGAGGAAGATCTCGCCTTTTGTCTCATTATGCTGTCCAGA
GACAAATGGAAGCAACAGAAGAAGAAGCAACGTGTAGAAGAAGATGAGACAGATCAT
GACAGTGAAGATTACAAATCAAGCAAGAGCAGAGGGAGATTCAAGTGTGAGACTTGTGGT
AAAGTGTTTAAATCTATCAAGCATTAGGAGGACACAGAGCAAGCCACAAGAAGAACAAG
GCATGCATGACGAAAACAGAGCAAGTTGAAACAGAGTACGTTCTTGAGTAAAGGAGAAG
AAAGTTCATGAATGTCCGATCTGTTTTAGGGTTTTTACTTCAGGGCAAGCACTTGGAGGT
CATAAGAGATCTCACGGAAGTAACATCGGAGCAGGAAGAGGATTGTCTAGTAAGTCAAATT
GTCCAAATCGAAGAAGAAGTATCAGTGAAACAGAGGATGATTGATCTTAATCTTCCTGCA
CCTAATGAAGAAGATGAAACTTCTTTGGTGTGTTGATGAATGGTGA
>G1914 Amino Acid Sequence (domain in AA coordinates:195-216, 245-266)
MERYKCRPFCFKSFINGRALGGHMRSHMLTLSAERCVITGEAEVEEERPSQLCDDDDDE
SDASSSSGEFDNQMNRLDDELEFDFAEDDDVESETESSRINPTRRRSKRTRKLSGDFD
FEKLTSQPSSELVAEPEHHSSASDTTTEEDLAFCLIMLSRDKWKQKQKKKQKQVEEDET
DSEDYKSSKSRGRFCKETCGKVFKSYQALGGHRAHSHKKNKACMTKTEQVETEVVLGVKEK
KVHECPICFRVFTSGQALGGHKRSHGNSNIGAGRLSVSQIVQIEEVSVKQRMIDLNLPA
PNEEDETSLVFDEW*
>G1925 (1..945)
ATGGAAGAAAATCTTCTCCGGGGTTCAGATTTTCATCCTACAGACGAGGAGCTCATAACG
CATTATCTATGTCCGAAAGTCTCCGATATAGGATTCACCGGTAAAGCTGTCTGACGTT
GATCTCAACAAGTGTGAACCTTGGGATTTGCCAGCCAAGGCTTCAATGGGAGAGAAAGAG
TGGTATTTCTTCAGCCAAAGGGATCGGAAATATCCAACCGGTTTAAAGAACAAACCGGCA
ACAGAAGCTGGTTACTGGAAAACACCGGGAAGATAAAGAAATATACCGAAGTGGAGTG
TTGGTTGGGATGAAGAAAACCTAGTTTCTACAAAGGAAGAGCTCCCAAAGGTGAGAAA
AGCAATTGGGTTATGCATGAGTACAGGCTTGAGAGCAAACAACCTTCAACCCACGAAT
AAGGAGGAATGGGTAGTGTGTAGGGTTTTCGAAAAGAGCACGGCAGCAAAGAAAGCACA
GAACAACAACCTCAATCTTCTCAACCATCTTTTGGATCTCCATGCGATGCAAACCTCATCA
ATGGCAAATGAGTTTGAAGATATTGATGAGCTTCCGAATCTGAATTCAAACCTCATCAACC
ATCGATTACAATAATCATATCCATCAATATTTCGCAACGCAATGTTTACTCAGAAGACAAC
ACAACAAGTACGGCTGGTCTCAACATGAACATGAACATGGCTAGTACTAATCTTCAGTCT
TGGACAACAAGTCTCCTTGGTCCGCCTTTATCTCCAATCAACTCTTTGTTGCTCAAGGCT
TTCCAAATCAGGAACCTTATAGTTTCCCAAAGAGATGATCCCAAGTTTCAATCATTTCT
TCTCTTCAACAAGGAGTCTCCAATATGATCCAAAATGCTTCAAGTTCGTCTCAAGTGCAA
CCCCAACCGCAAGAGGAAGCGTTTAAATATGGACTCCATATGGTGA
>G1925 Amino Acid Sequence (conserved domain in AA coordinates:6-150)
MEENLPPGFRFHPTDEELITHYLCRKVSDIGFTGKAVVDVLDNKCEPWLPAKASMGEKE
WYFFSQDRDKYPTGLR TNRATEAGYWKTTGDKDEIYRSGVLVGMKKT LVFYKGRAPKGEK
SNWVMHEYRLSKQPFNPNTNKEEWVVC RVFEKSTAAKKAQEQQPQSSQPSFGSPCDANSS
MANEFEDIDELPNLNSNSTIDYNNHIHQYSQRNVYSEDNTTSTAGLNMNMNMMASTNLQS
WTTSLGPPPLSPINSLLLKAFQIRNSYSFPKEMIPSFNHSSLQQGVSNMIQNASSSSQVQ
PQPQEEAFNMDSIW*
>G1929 (1..366)
ATGTGTAGAGGCTTGAATAATGAAGAGAGCAGAAGAAGTGACGGAGGAGGTTGCCGGAGT
CTCTGCACGAGACCGAGTGTTCGGTAAGGTGTGAGCTTTGCGACGGAGACGCCTCCGTG
TTCTGTGAAGCGGACTCGCGCTTCTCTGTAGAAAATGTGACCGGTGGGTTTCATGGAGCG
AATTTTCTAGCTTGGAGACAGTAAGGCGCGTGTGCTATGCACTTCTGTGAGAACTCAGC
CGCCGGTGCGCTCGTGGAGATCATGACTTCCACGTTGTTTTACCGTCCGTGACGACGGTC
GGAGAAACCACCGTGGAGAATAGAAGTGAACAAGATAATCATGAGGTTCCGTTTGTGTTTT
CTCTGA
>G1929 Amino Acid Sequence (domain in AA coordinates:31-53)
MCRGLNNEESRRSDGGGCRSLCTRPSVPVRCELCDGDASVFCEADSAFLCRKCDRWVHGA
NFLAWRHVRRVLCTSCQKLTRRCLVGDHDFHVVLPSVTTVGETTVENRSEQDNHEVPFVF

L*

>G1930 (76..1077)

ATTACATTACTAATCTCTCAAGATTTTCAAAATTTTCTTGATTTTCTCTCAGTTTCTT
ATTTTCGTTTTCATAACATGGATGCCATGAGTAGCGTAGACGAGAGCTCTACAACACAGAT
TCCATTCCGGCGAGAAAGTATCGTCTCCGGCGAGTTTACTATATAGAATGGGAAGCGGA
ACAAGCGTGGTACTTGATTTCAGAGAACGGTGTGCAAGTCAAGTCAAGCCGAATCAAGA
AAGCTTCTTCTTCAAGATTCAAAGGTGTGTTCTTCAACCAAATGGAAGATGGGGAGCT
CAGATTTACGAGAAACATCAACGCGTGTGGCTTGGTACTTCAACGAGGAAGACGAAGCA
GCTCGTGCTTACGACGTCGCGGCTCACCGTTTCCGTGGCCGCGATGCCGTTACTAATTTT
AAAGACACGACGTTTCAAGAAGAGGTTGAGTTCTTAAACGCGCATTTCGAAATCAGAGATC
GTAGATATGTTGAGAAAACACACTTACAAGAAGAGTTAGACCAAAGGAAACGTAACCGT
GACGGTAACGGGAAAAGAGACGACGCGGCTTTGCTTTGGCTTCGATGGTGGTTATGACGGGG
TTTTAAACCGGCGAGTTACTGTTTGAAGAAACGGTAACGCCAAGTGACGTCGGGAAACTA
AACCCTTTAGTTTATACAAAACACCAAGCGGAGAAACATTTCCGTTACCGTTAGGTAAT
AATAACGTCTCCGTTAAAGGTATGCTGTTGAATTTCAAGACGTTAACGGGAAAGTGTGG
AGGTTCCGTTACTCTTATTGGAATAGTAGTCAAAGTTATGTTGACCAAAGGTTGGAGT
AGATTCGTTAAAGAGAAGAGACTTTGTGCTGGTGGTATGATCAGTTTAAAGATCCAAC
GATCAAGATCAAAAATCTTTATCGGGTGGAAATCGAAATCCGGGTTGGATCTAGAGACG
GGTCGGGTTATGAGATTGTTTGGGGTTGATATTCTTTAAACGCCGTCGTTGTAGTGAAG
GAAACAACGGAGGTGTTAATGTCGTCGTTAAGGTGTAAGAAGCAACGAGTTTGTATAAA
CAATTTAACAACTTGGGAAAGAAAAAAGCTTTTGTATTTAATTTCTCTTCAACGTTA
ATCTTGCTGAGATTA

>G1930 Amino Acid Sequence (domain in AA coordinates: 59-124)

MDAMSSVDESSTTTDSIPARKSSSPASLLYRMGSGTSVVLDSENGVEVEVEAESRKLPS
RFKGVVPPQNGRWGAQIYBKHQVWLGTFFNEEDEAARAYDVAHRFRGRDAVTNFKDTTF
EEVEFLNAHSKSEIVDMLRKHTYKEELDQKRNRDNGKETTFALASMVVMTGFKTAE
LLFEKTVTPSDVGLNRLVIPKHQAEKHFPLPLGNMNVSVKGMMLNFEDVNGKVWRFYS
YWNSSQSYVLTKGWSRFVKEKRLCAGDLISFKRSNDQDQKFFIGWKS KSGLDLETGRVMR
LFGVDISLNAV VVVKETTEVLMSLRCKKQVRV*

>G195 (51..1031)

TTTTCTTTTCTTTCTTTTGGTTTAAAGTTTTTCTCTTTGTTCTTCGTCATGTCTCATG
AAATCAAAGATCTTAACAACTACTACTACATTCATCGTATAATCATTACAATATCAACA
ACCAAATATAGATTAATCTCCCTTACGTTTCTGGTCCATCTGCTTATAATGCAACATGA
TCTCATCATCACAAGTAGGTTTGTATCTACCTCGAAGAACTTGAGTCCTCAAGGAGCCT
TCGAGTTGGGTTTCGAGCTTTCTCCATCTTCTTGACTTTTTTAATCCTTCCCTCGATC
AAGAGAACGGTTTGTATAATGCTTATAATATAATAGTAGTCAAAAGAGTCATGAAGTTG
TCGGTGATGTTGTGCAACCATTAAAGAGTGAAGTTAGGGTTTCAGCATCTCCTTCTTCAA
GTGAGGCCGATCATCATCCAGGAGAAGATTCGGCAAGATCCGGAAGAAAAGAGAAGTTC
GCGATGGAGGAGAAGATGATCAACGCTCTCAGAAAGTAGTTAAAACAAAGAAGAAAGAGG
AGAAGAAAAAAGAGCCACGAGTCTCGTTTCATGACTAAGACCGAAGTTGATCATCTCGAAG
ACGGCTATCGTTGGAGAAAGTATGGCCAAAAAGCAGTCAAAAACAGTCCTTATCCGAGGA
GTTACTATAGATGCACGACTCAGAAGTGCAACGTGAAGAAGAGAGTGGAGAGATCTTACC
AAGACCCAACGGTTCGTCATCACAACCTACGAGAGTCAACACAACCATCCGATCCCGACCA
ATCGTCGGACAGCAATGTTCTCTGGAACCCCGCATCTGATTATAACCCATCATCGTCTC
CAATATTCTCCGATCTCATCATCAATACTCCAAGAAGCTTCTCAAATGATGATCTCTTCC
GTGTGCCATACGCTAGTGTGAACGTGAACCTAGTTATCATCAACAGCAACATGGATTTC
ATCAACAGGAGAGTGAGTTTCGAGCTCTTGAAGGAGATGTTTCCTTCGGTTTTCTTCAAAC
AAGAGCCTTGATGATATAATATAATATAGAAACAATTTTTTTCTGCTAAGAAATATAGA
ACAAAACCTGGATGCATAATAAGTGATGATAGTGTTATTTATTTTTTGCATGTATATATT
ATACATGTTTTGTTAACTAGCTATAGGATATACTGGTAGTAATTAAGCATAAATATGGAG
CCCTTCGACTTATTACAATAATTTTGGTATGAAAAANTTNGNTACATGCCTGCCTTTT
NNNTTNNNG

>G195 Amino Acid Sequence (domain in AA coordinates: 183-239)

MSHEIKDLNNYHYTSSYNHYNNQNMNINLPYVSGPSAYNANMISSSQVGFDLPSKNLSP
QGAFELGFELSPSSSDFFNPSLDQENGLYNAYNINSSQKSHEVVGDGCATIKSEVRVSAS
PSSSEADHHPGEDSGKIRKKREVRDGGEDDQRSQKVVKTKKKEKKKEPRVSFMTKTEVD
HLEDGYRWRKYGQKAVKNSPYPRSYRCTTQKCNVKKRVERSYQDPTVVITYESQHNHP

IPTNRRRTAMFSGTTASDYNPSSSPIFSDLIINTPRSFSDNDLFRVPYASVNVNPSYHQQQ
HGFHQESEFELLKEMFSPVFFKQEP*

>G1954 (196..1440)

ATTTATGACTTCTCAATACAAAAAGCTCCCTCACTTTTTTAAGTTTGTCTTCTCTAAT
CCGTCTTCTTCTACTATCTTGCATGTCTTGGCTCTTTATATACATCTCTCGTAAACCCT
AGCAAATCATACAAGGTCAAGAAGCTTGACCTTCATTAGACTTAAGCAGTTTATAATCAA
CTACCACGAATAGCAATGGATAAAGATTACTCGGCACCAAACCTTCTTAGGTGAATCCTCA
GGCGGTAAACGATGATAACAGCTCTGGTATGATAGACTATATGTTCAATAGAAACCTTCAA
CAACAACAAAAGCAATCGATGCCACAACAGCAGCAACATCAACTCTCTCCTTCCGGATTT
GGAGCAACACCCTTTTGATAAAATGAAGCTTCTCTGATGTGATGCAGTTTGCGGACTTCGGT
TCGAAACTTGCGTTGAACAGACCAGAAACCAAGACGATCAAGAAACCGGGATTGACCC
GTTTATTTCTTGAAGTTCCTGTCTTGAACGACAAAATAGAGGACCATAACCAAACCCAA
CATCTCATGCTTCTCATCAGACGTCTCAAGAAGGAGGTGAGTGTGGAGGAAACATAGGC
AATGTGTTTCTTGAAGAAAAAGAAGATCAAGACGATGACAACGACAACAACCTCCGTGCAA
CTACGTTTTATTGGAGGAGAAGAAGATAGGGAGAACAAAGATGTTACGAAAAAGGAG
GTGAAGAGCAAGAGGAGAGAGCTAGAACGAGCAAGACCAGCGAAGAAGTGGAAAGCCAA
CGGATGACTCATATCGCGTTCGAAAGAAACCGTAGGAGCAAAATGAATGAGCATCTTCGT
GTCCTTAGATCTCTCATGCTGGCTCCTACGTTCAAAGGGGAGACCAAGCGTCAATCATA
GGAGGAGCAATAGAGTTTGTGAGAGAGCTCGAGCAACTCCTACAATGTCTTGAATCACAG
AAGCGTCGAAGAATCTTAGGAGAAACCGTAGGGACATGACAACGACAACGACTTCTTCT
TCTTCTCCCATAACTACGGTAGCGAACCAAGCACAACCGCTCATTATTACGGGAAATGTA
ACCGAGCTAGAGGGCGGAGGAGGGCTTCGGGAGGAGACTGCGGAGAACAAAGTCGTGCTTG
GCTGACGTGGAGGTGAAGCTGCTAGGGTTTGACGCCATGATCAAGATACTTTCAAGAAGA
AGGCCGGGACAGCTGATTAAAGACTATAGCTGCTTTGGAGGATCTTCATCTCTCTATTCTT
CACACTAACATCACTACCATGGAACAAACCGTCTCTACTCTTTAATGTCAAGATAACA
AGTGAACAGAGGTTTACGGCAGAAGACATAGCAAGTTCATCCAACAGATATTTAGTTTC
ATTTCATGCAAAATACCAACATATCTGGAAGCTCTAACCTGGGAAATATTGTGTTTACTTGA
AAATCATCACACGGCGACAACCTTTGTACACTGGTGAAGATTACAGTACGTAATAATCTCT
ACATATTGGGTTTTATTCTCCAAGCATTTGGAAGAGTGTAAAGTTAAAGGGAGTGCTTA
CTTTATTTTTTTGGGGCTTTTTTCATGCAATTTAAATTTTAGTGTGATTGTGTCGCTTG
TAATGTTAGAAGCTCGTTGTTGTGATTCTGCTGCTTTGATTGTAGGTTTTGAACAAGCG
GTTTAGAATGCTAAACCCTTATTACTTGAAATAACTTTTTTCACAAAAA

>G1954 Amino Acid Sequence (domain in AA coordinates:187-259)

MDKDYSAPNPLGESSGNDNDSGIMIDYMFNRNLQQQQKQSMPPQQQHQQLSPSGFGATPF
DKMNFSDVMQFADFSGKLALNQTRNQDDQETGIDPVYFLKFPVLNDKIEDHNQTHLMPS
HQTSQEGGECGNIGNVLEEKEDQDDNDNNSVQLRFIGEEEDRENKKNVTKEVKSKR
KRARTSKTSEEVESQRMTHIAVERNRRKQMNHLRVLRLSLMPGSYVQRGDQASIIGGAIE
FVRELEQLQLCLESQKRRRILGETGRDMTTTTSSSPITTVANQAQPLIITGNVTELEG
GGGLREETAENKSLADVEVKLLGFDAMIKILSRRRPGQLIKTIAALEDLHLSILHTNIT
TMEQTVLYSFNVKITSETRFTAEDIASSIQQIFSFHANTNISGSSNLGNIVFT*

>G1958 (107..1336)

GTACCGTCGACCGATTATCCCCAAGAGGAGAATCCTCATAATCATTTTCTCCGATTTCGAT
TCGTCTTCCCTGGTCCCTGGATTGCTTCATGAATTTCTAGGACAACAATGGAGGCTCGTCC
AGTTCATAGATCAGGTTTCGAGAGACCTCACACGCACTTCTTCAATCCCATCTACACAAAA
ACCTTCACCAAGTAGAAGATAGTTTCATGAGATCAGATAACAACAGTCAGTTAATGTCTAG
ACCATTAGGACAAACCTACCATTTACTTTCATCTAGTAACGGTGGAGCTGTTGGACATAT
ATGTTCTTCTTCATCATCTGGTTTTGCAACCAATCTCCATTACTCAACTATGGTATCTCA
TGAGAAACAACAACACTACACAGGAAGCAGCAGTAATAATGCTGTGTCAGACACCAAGCAA
CAACGATAGTGCTTGGTGTGATGATTCATTGCCAGGAGGGTTTCTGACTTCCATGAAAC
CAACCCGGCGATTCAAAACAACGTGCAGATTGAGGATGGTGGCATTGCGGCTGCTTTGA
TGACATTCAAAACGAAGTATTGGCATGAATGGGCTGACCATTGATCACTGATGATGA
TCCTTTGATGTCTACTAAGTGAATGATCTCTTGCTTGAAACAAATCCAATTGATTC
AAAGGACCAGAAGACTGCAAATTCGCAACCTCAGATTGTTTCAGCAGCAACCTTCTCC
GTCTGTGGAATTGCGACCTGTTAGCACACATCTTCAAACAGCAATAACGGAACGGGCAA
GGCAGCAATGCGTTGGACGCCAGAGCTTCACGAGGCTTTGTTGAGGCTGTCAACAGTCT
TGGCGGTAGTGAAAGAGCTACTCTAAAGGGTACTGAAGATTATGAAAGTTGAAGGCTT

GACTATATATCATGTAAAGCCATTTACAGAAATATAGGACAGCTAGATATCGGCCAGA
 ACCATCAGAACTGGTTCGCCAGAAAGGAAGTTGACACCGCTTGAACATATAACATCTCT
 TGATTTGAAAGGTGGGATAGGTATTACAGAGGCTCTACGACTTCAGATGGAAGTACAGAA
 GCAACTCCATGAGCAGCTCGAGATTCAAAGAAACCTGCAACTCCGAATAGAAGAACAAAGG
 CAAGTACCTGCAATGATGTTTCGAGAAGCAAACTCTGGTCTTACCAAAGGGACAGCCTC
 AACATCAGATTCGCGAGCCAAATCTGAACAAGAAGACAAGAAGACTGCTGATTCTGAAGGA
 GGTTCCAGAAGAAGAAACCAGGAAATGTGAGGAAGTAGAATCTCCACAGCCAAAGCGTCC
 CAAAATCGATAATTGAAAGTATTGGTCTTTTGTGCTGATAATCTCGAGTTTCAGAGTTAA
 CAGTGATAGAGAGAACGAGCTCTTATCTTGAGGTTCTTCAGGACTTCTCTCGCGGCCGCT
 CTAG

>G1958 Amino Acid Sequence (domain in AA coordinates: 230-278)

MEARPVHRSGSRDLTRTSSIPSTQKPSPVEDSFMRSDMNSQLMSRPLGQTYHLLSSSNGG
 AVGHICSSSSSGFATNLHYSTMVSHEKQQHYTGSSSNNAVQTPSNNSAWCHDSLPGGFL
 DFHETNPAIQNNCQIEDGGIAAFDDIQKRSDWHEWADHLITDDPLMSTNWNLDLLETN
 SNSDSKDQKTLQIPQPQIVQQPSPVELRPVSTSSNSNNGTGKARMRWTPELHEAFVE
 AVNSLGGSERATPKGVLMKIMVEGLTIYHVKSHLQKYRTARYRPEPSETGSPERKLTPL
 HITSLDLKGGIGITEALRLQMEVQKQLHEQLEIQRLQLRIEEQGYLQMMFEKQNSGLT
 KGTASTSDSAKSEQEDKKTADSKEVPEETRKCCELESPPQKRPKIDN*

>G196 (111..1421)

TCGACATCAGATTTCTCTACGGATTCTTAATCATTTTTATTATATTGGATATTTGCTA
 ATTCTTCCCGTGTATAAATCTCATATAAACACGCATCATACATATATATTATGTGCAGCG
 TCTTTGAGTTTCAAGACATGGACAACCTCCAAGGAGATCTAACAGACGTCGTACGAGGAA
 TAGGATCAGGCCACGTGTCACCATCTCCTGGACCACCGGAAGGTCCATCTCCGAGCAGCA
 TGTCTCCGCCGCCAACATCAGATCTCCACGTGGAATTCCTCCGCCGCTACTTCTGCCA
 GCTGTCTCGCAAATCCCTTCGGAGACCCGTTCTGTAAGCATGAAGGATCCTCTCATCCACC
 TCCCGGCCAGCTACATCTCCCGGCCCGGTGATAATAAAGCAACAAAAGTTTTGCAATCT
 TTCCAAAGATTTTGTAGGATGATCATATTAAGAGTCAATGCAGTGTCTTCCCAAGAATTA
 AGATCTCGCAAAGTAACAATATCCACGATGCCTCCACGTGTAATTCTCCGCCATAACCG
 TCTCCTCTGCCGCCGTAGCAGCTTCGCCGTGGGGCATGATCAACGTTAATAACCACTAACA
 GTCCAAGAACTGTTTACTTGTGCGATAATAAACAACACGTCATCATGCTCACAGGTTTC
 AGATCTCTTCTTCCCTCGGAATCTCGGAATTAAGAGAAGGAAGAGCCAGGCAAAGAAAG
 TGGTGTGCATACCGGCTCCAGCCGCTATGAACAGCCGTCAGTGGAGAAGTTGTTCCGT
 CTGATCTATGGGCTTGGCGAAAGTACGGTCAAAAACCTATCAAAGGTTCTCCTTATCCAA
 GGGGTTACTACAGATGTAGCAGCTCAAAGGTTGTTTCAGCTAGGAAACAAGTCGAACGTA
 GCCGCACTGATCCAAACATGTTAGTCATTACTTACACCTCTGAGCATAACCAACCCATGGC
 CTACTCAACGCAACGCTCTCGCAGGTTCCACTCGTTCCTCTTCTCCTCCTCTTTAAACC
 CTTCTTCCAAATCCTCAACCGCAGCCGCCACTACTTCTCCCTCATCCAGATTTTCCAAA
 ACAACAGCAGCAAAGACGAACCCAATAACTCCAATTGCCTTCTCTTCCACTCATCCTC
 CTTTGTAGCGCCGCCGAATTAAGGAGGAGAACGTGGAAGAGCGTCAGGAAAAGATGGAGT
 TCGATTATAATGACGTTGAAAATACCTATAGACCGGAGTTGTTGCAAGAGTTTCAACATC
 AGCCGGAGGATTTCTTTGCCGATCTCGACGAGCTTGAGGGAGATTCTTTGACTATGTTGC
 TCTCTCACAGTAGCGCGGAGGCAACATGGAAAACAAAACGACGATTCCAGACGTTTTTA
 GTGATTTCTTTGACGACGACGAGTCCTCAAGGTGTTATAAATATTGTTGTTAATGTATA
 CATAGAAATGAAATTATTCATGTAATTCGTTTTGTGTTAAATGACGGTATTTGCCTTTC
 A

>G196 Amino Acid Sequence (conserved domain in AA coordinates: 223-283)

MCSVFEFQMDNMFQGDLDVVRGIGSGHVSPSPGPPEGSPSPSSMSPPPTSDLHVEFPSAA
 TSASCLANPFGDPFVSMKDLIHLPAASYISGAGDNKSNKSFAPFKIFEDDHISQCSVF
 PRIKISQSNNIHDASTCNPAITVSSAAVAASPWGMINVNTTNSPRNCLLDVNNNTSSC
 SQVQISSSPRLGIKRRKSQAKKVVCIPAPAAMNSRSSGEVVPSDLWAWRKYQKPIKGS
 PYPGRYYRCSKSGCSARKQVERSRTDPNMLVITYTSEHNHPWPTQRNALAGSTRSSSSS
 SLNPSSKSSATAATTSPSSRVFQNNSSKDEPNNSNLPSSTHPPFDAAAIKEENVEERQE
 KMEFDYNDVENTYRPELLQEFQHQPEDFFADLDELEGDSLTMLLSHSSGGGNMENKTTIP
 DVFSDFDDDESSRSL*

>G1965 (1..609)

ATGGATAACTTCAATGTTGTTGCCAATGAAGACAATCAAGTGAATGATGTGAAGCCTCCA
 CCACCCACCACCGAGTGTGTGCAAGATGTGATCTGATAACACAAAATTTGTTACTAC

AACAATTATAGTGAGTTTCAACCGCGCTACTTCTGCAAGAACTGTGGAAGATACTGGACT
CATGGTGGGGCTTTAAGAAACGTACCAATTGGTGGGAGTAGTCGTGCCAAGCGGACAAGG
ATAAATCAACCTTCAGTTGCTCAGATGGTTTCTGTTGGAATCCAACCAGGGAACCGTTT
AGTTCTTTGTCTCATATTCATGGTGGTATGGTAACAAATGTGCATCCAACCTCAAACCTTT
CGACCAAATCATCGCCTAGCTTTCCATAATGGATCATTTGAGCAAGATTATTATGATGTT
GGGTCTGATAATCTTTTGGTAAACCAACAGTTGGTGGATATGTTGATAATCACAACGGT
TATCACATGAATCAAGTGGATCAATACAACCTGGAACCAGAGCTTCAATAACGCTATGAAC
ATGAATTATAATAACGCTAGCACTAGCGGAAGGATGCATCCTAGTCATTTAGAGAAGGGT
GGTCCTTGA

>G1965 Amino Acid Sequence (domain in AA coordinates:27-55)
MDNFNVVANEDNQVNDVKPPPPPRVCARCDSDNTKFCYNNYSEFQPRYFCNKNCRRYWT
HGGALRNVPIGSSRAKRTRINQPSVAQMVSVGIQPGNRFSSLSHIHGGMVTNVHPTQTF
RPNHRLAFHNHGSFEQDYDVGSDNLLVNQQVGGYVDNHNHGYHMNQVDQYNWNQSFNNAMN
MNYNNASTSGRMHPSHLEKGGP*

>G1976 (1..1152)
ATGACTGATCCTTATTCCAATTTCTTCACAGACTGGTTCAAGTCTAATCCTTTTCACCAT
TACCCTAATTCCTCCACTAACCCCTCTCCTCATCCTCTCCTCCTGTACTCCTCCCTCT
TCCTTCTTCTTCTTCCCTCAATCCGGAGACCTCCGCCGTCCACCGCCGCCACCAACTCCT
CCTCCTTCTCCTCCTCTCCGAGAAGCCCTCCCTCTCCTCAGCCTCAGCCCCGCCAACAAA
CAACAAGACCACCATCACACCATGACCACCTTATTCAAGAACCACCTTCAACCTCCATG
GATGTCGACTACGATCATCACCATCAAGATGATCATATAACCTCGATGACGATGACCAT
GACGTCACCGTTGCTCTTCACATAGGCCCTTCAAGCCCTAGTGCTCAAGAGATGGCCTCT
TTGCTCATGATGCTTCTTCTTCTCTCTCCTCGAGGACCACTCATCATCAGAGGACATG
AATCACAAGAAAGACCTCGACCATGAGTACAGCCACGGAGCTGTGCGAGGAGGAGAAGAT
GACGATGAAGATTAGTCGGCGGAGACGGCGGCTGTAGAATCAGCAGACTCAACAAGGGT
CAATATTGGATCCCTACACCTTCTCAGATTCTCATTGGCCCTACTCAGTTCTCATGTCCT
GTTTGCTTCAAAACCTTCAACAGATACAATAACATGCAGATGCATATGTGGGGACATGGA
TCACAATACAGAAAAGGACCTGAATCTCTAAGGGGAACACAACCAACAGGAATGCTAAGG
CTTCCGTGCTATTGCTGCGCCCCAGGCTGTCGCAACAACATTGACCATCCAAGGGCAAAG
CCTCTCAAGACTTCAGAACCTTCAAACACATTACAAGAGAAAACATGGGATCAAACCT
TTCATGTGTAGAAATGTGAAAAGGCTTTCGCACTCCGAGGGGACTGGAGAACACATGAG
AAGAATTGTGGCAAACCTTGGTATTGCATATGTGGATCTGATTTCAAGCACAAGAGATCT
CTCAAAGATCACATCAAGGCTTTTGGGAATGGTCATGGAGCCTACGGAATTGATGGGTTT
GATGAAGAAGATGAGCCTGCCTCTGAGGTAGAACAATTAGACAATGATCATGAGTCAATG
CAGTCTAAATAG

>G1976 Amino Acid Sequence (domain in AA coordinates: 219-323)
MTDPYSNFFTDWFKSNPFHHYPNSSTNPSPHPLPPVTPPSSFFFFPQSGDLRRPPPPPTP
PPSPPLREALPLLSLSPANKQQDHHNHDHLIQEPPSTMDVDYDHHHQDDHNLDDDDH
DVTVALHIGLPSPSAQEMASLLMMSSSSSSSRTHHHEDMNHKKDLDEYSHGAVGGGED
DDEDSVGGDGGCRISRLNKGQYWIPTPSQILIGPTQFSCPVCFKTFNRYNNMQMHMWGHG
SQYRKGPESLRGTQPTGMLRLPCYCCAPGCRNNIDHPRAKPLKDFRTLQTHYKRKHGIKP
FMCRCCKGKAFVRGDRWTHEKNCGLWYICGSDFKHKRSLKDHIAFGNGHGHAYGIDGF
DEEDEPASEVEQLDNDHESMQSK*

>G2057 (27..1289)
GCCGTCTCGACGAATATGCTTACCAATGTCTGACGACCAATTCCATCACCCGCCGCCCTC
CTTCTTCAATGAGGCACCGTTCTACGTCGGATGCGGCGGACGGCGGCTGCGGCGAGATTG
TTGAGGTGCAAGGTGGTCACATTGTTCCGTCTACCGGAAGAAAAGACCGCCACAGCAAAG
TCTGCACGGCTAAAGGGCCACGTGACCGGCGCGTGAGACTCTCTGCTCACACGGCGATT
AGTTTTACGATGTTCAAGACAGGCTTGGTTTCGACCGACCTAGCAAAGCCGTTGATTGGC
TTATCAAAAAGGCTAAGACTTCCATTGACGAGCTCGCTGAGCTTCCCTCCCTGGAATCCCG
CCGATGCAATTCCGCTAGCGCTGCTAACGCTAAACCCAGAAGAACCACCGCCAAAACCC
AAATCTCTCCGTCTCCGCCACCGCGCAACAGCAACAACAACAACAGCTTCAGTTTCG
GTGTTGGCTTCAACGGAGGAGGAGCAGAGCATCCGAGTAACAACGAGTCGAGTTTCTCC
CGCCGTCAATGATTGAGATTCGATAGCTGACACTATAAAGTCGTTTTTCCGGTGATTG
GCTCTTCAACGAGGCTCCTTCGAATCATAACCTTATGCACAACATCATCATCAGCATC
CGCCGGATTGCTTTCTCGAATAATAGCCAAAACCAAGATCTCCGTCTCTCGCTGCAAT
CGTTCCCGGATGGTCCACCGTCGCTTCTGCACCACCAACATCACCACCACACCTCTGCTT

CCGCCTCCGAGCCTACTCTGTTCTACGGACAGAGCAATCCGTTAGGGTTTGACACATCGA
GTTGGGAGCAGCAGTCGTCGGAATTCGGAAGGATTAGAGACTAGTGGCTTGAACAGCG
GCGGTGGCGGCGGAGCAACCGATACAGGAAACGGAGGAGGGTTCTGTTGCTCCTCCTA
CTCCTTCAACGACGTCGTTTACGCCAGTTCTTGCCAAAGCCAACAGCTTTATCTCAGA
GGGGTCCCCTTCAGTCCAGTACAGTCCCATGATCCGTGCTTGGTTGATCCTCACCATC
ATCACCATTCATCTCCACCGACGATCTCAACCACCACCATCACCTTCTCCACCGGTTT
ACCAATCAGCAATCCCCGAATCGGATTCGCCTCAGGTGAATTCTTTCGGGTTTTCGCA
TACCAGCAGGTTTTCAGGGCCAAGAAGAGGAGCAGCAGCAGGTTCTCACTACAAGCCGT
CCTCTGCTTCTCTATTCTCGCCATTGACAATCGAACTAATCCTC

>G2057 Amino Acid Sequence (domain in AA coordinates: TBD)

MSDDQFHHPPPPSSMRHRSTSDAADGGCGEIVEVQGGHIVRSTGRKDRHSKVC TAKGPRD
RRVRLSAHTAIQFYDVQDRLGFDRPSKAVDWLIKAKTSIDELAE LPPWNPADAIRLAAA
NAKPRRTTAKTQISPSPPPPQQQQQQQLQFGVGFNGGGAEHPSNNESSFLPPSMDSDSI
ADTIKSFPPVIGSSTEAPSNHNLHNYHHQHPPDLLSRTNSQNQDLRLSLQSPDGPPL
LHHQHHTSASASEPTLFYQSNPLGFDTSWEQSSSEFGRIQRLVAWNSGGGGGATDT
GNGGGFLFAPPTPSTTSFQVLGQSQQLYSQRGPLQSSYSPIRAWFDPHHHQSISTDD
LNHHHHLPPP VHQSAIPGIGFASGEFSSGFRIPARFQGEQEHDLTHKPSSASSISRH

*

>G2107 (79..624)

ACCACAAAACAGAGCAACACACAACACAAAGCTTCATTTCAATTCTGTTTCGAGAACCCT
TTGAGAACCAGATCGGAGATGGAACGACGATATCACCGTGGCGGAGATGAAGCCAAAG
AAGCGTGCTGGACGAGGATTTCAAGGAGACACGTACCCAATCTACAGAGCGTGCGG
CGTAGGGACGGCGACAAATGGGTATGCGAAGTCCGTGAACCGATTATCAGCGTCGAGTC
TGGCTCGGAACTTATCCGACGGCAGATATGCCCGCACGTGCTCACGACGTGGCGGTTCTT
GCTCTGCGCGGGAGATCCGCGTGTGTAATTTCTCCGATTCTGCTTGGAGGTTGCCGGTG
CCGGCATCCACTGATCCGGACACGATCAGGCGCACGGCGCCGAAGCAGCGGAGATGTTT
AGGCCGCGGAGTTTAGTACAGGAATTACGGTTTACCCTCAGCCAGTGAGTTTGACACG
TCGGATGAAGGAGTCGCTGGAATGATGATGAGGCTCGCGGAGGAGCCGTTGATGTCGCCG
CCAAGATCGTACATTGATATGAATACGAGTGTGTACGTGGACGAAGAAATGTGTTACGAA
GATTTGTCACTTTGAGTTACTAAAATACGTATGTGTTAAAAAACCAAGATCGTATGTG
TATGTATGCATAATAAATGGGCTTAATGATGGGCATAGATATGATAGGTCCAGCCTATAT
GTTAAATGTGTTTTATTTTTGGTTTATCTAGTTTCTAGGTATTTACCAATTGTATTA
GTATAAGTTTATTATAAGAAATAATCAAAAATGTTGTTGCCAAAAA AAAAAAAAAA
AAAAA

>G2107 Amino Acid Sequence (domain in AA coordinates: TBD)

MENDDITVAEMKPKKRAGRRIFKETRHPYIRGVRRRDGDKWVCEVREPIHQRRVWLGTYP,
TADMAARAHDVAVLALRGRSACLNFSDSA WRLPVPASTDPDTIRRTAAEAEMFRPPEFS
TGITVLPSASEFDTSDG VAGMMMLAEPELMSPPRSYIDMNTSVYVDEEMCYEDLSLWS
Y*

>G211 (1..750)

ATGATGTCATGTGGTGGGAAGAAGCCAGTGCTAAGAAAACAACGCCGTGTTGCACGAAG
ATGGGGATGAAGAGAGGACCATGGACGGTGGAGGAAGACGAGATTCTTGTGAGCTTCATT
AAGAAAGAAGGTGAAGGACGGTGGCGATCGCTTCTAAGAGAGCTGGTTTACTCAGATGT
GGAAAGAGCTGTGCTCTACGGTGGATGAACTATCTCCGACCCCTCGGTTAAACGTGGAGGA
ATTACGTCCGACGAGGAAGATCTCATCTCCGTCTTACCGCCTCCTCGGCAACAGGTGG
TCATTGATCGCGGAAGGATACCGGGAAGGACTGATAATGAAATTAAGAATAATTGGAAC
ACTCATCTTCGTAAGAACTTTTAAGGCAAGGAATTGATCCTCAAACCCACAAGCCTCTT
GATGCAACCAACATCAAAACAGAGAAGAAGTTTCCGGTGGACAAAAGTACCCTCTA
GAGCCTATTTCTAGTTCTCATACTGATGATACCACTGTTAATGGCGGGGATGGAGATAGC
AAGAACAGTATCAATGTCTTTGGTGGTGAACACGGCTACGAAGACTTTGGTTTCTGCTAC
GACGACAAGTTCTCATCGTTTCTTAATTCGCTCATCAACGATGTTGGTGATCCTTTTGGT
AATATTATCCCAATATCTCAACCTTTGCAGATGGATGATTGTAAGGATGGGATTGTTGGA
GCGTCGTCTTCTAGCTTAGGACATGACTAG

>G211 Amino Acid Sequence (conserved domain in AA coordinates: 24-137)

MMSGGKKPVSKTTPCCTKMGMKRPWTVEEDELVSFIKKEGGRWRS LPKRAGLLRC
GKSCRLRWNNYLRPSVKRGGITSD EEDLILRLHRLGNRWSLIAGRIPGRTDNEIKNYWN
THLRKKLLRQGIDPQTHKPLDANNIHKPEEEVSGGQKYLEPISSHTDDTTVNGGDGDS

KNSINVFGEHGYEDFGFCYDDKFSSFLNSLINDVGDPFGNIIPISQPLQMDDCKDGIIVG
ASSSSLGHD*

>G2133 (26..457)

ATCTCATCTTCATCCACCCAAAAACATGGATTCAAGAGACACCGAGAACTGACCAGAG
CAAGTACAAAGGTATCCGTCGTCGGAATGGGGAATGGGTATCAGAGATTCGTGTCCC
GGGAACCTCGTCAACGTCTCTGGTTAGGCTCTTTCTCCACCGCAGAAGGCGTGCCGTAGC
CCACGACGTCGCTTTTACTGCTTGCACCGACCATCTTCCCTCGACGACGAATCTTTTAA
CTTCCCTCACTTACTTACAACCTCCCTCGCCTCCAATATATCTCCTAAGTCCATCCAAAA
AGCTGCTTCCGACGCCGGCATGGCCGTGGACGCCGGATTCCATGGTGCTGTGTCTGGGAG
TGGTGGTTGTGAAGAGAGATCTTCCATGGCGAATATGGAGGAGGAGACAACTTAGTAT
CTCCGTGTATGATTATCTTGAAGACGATCTCGTTTGATCTATACGAGTACGTTTGTAGCA
GTAA

>G2133 Amino Acid Sequence (domain in AA coordinates:11-83)

MDSRDTGETDQSKYKIGIRRRKWGWVSEIRVPGTRQRLWLGSFSTAEGAAVAHDVAFYCL
HRPSSLDDSEFNFPHLLTSLASNISPKSIQKAASDAGMAVDAGFHGAVSGSGGCEERS
MANMEEEDKLSISVYDLEDDL*

>G2134 (36..644)

GAGCAAAAACTTTGTGTGCGTGTGTGTGTGTTCATGGCTGGTCTTAGGAATTCGGTA
ACAGCGACAAAGCGCAAAACGATGGCAAAGGTGTACCATCTGCCTACAGAGGAGTCCGGA
AGAGAAAATGGGGGAAATGGGTGTCTGAAATCCGTGAACCGGGGACCAAGAACCGTATCT
GGCTAGGCAGTTTCGAGACTCTGAAATGGCTGCAACCGCATACGACGTGGCAGCATTTC
ATTTACAGAGGGAGAGAAGCTCGTCTCAACTTCCCTGAGCTCGCCAGCAGCCTTCCACGTC
CTGCAGACTCTAGCTCAGACAGCATTCGCATGGCAGTTTCATGAGGCAACACTCTGCCGCA
CCACCGAAGGAACAGAGTCAGCATGCAAGTGGACAGCTCAAGCTCCTCCAATGTAGCTC
CAACAATGGTCAGACTCTCGCCCAGGGAATCAAGCGATCAACGAGTCAACTTTGGGAT
CTCTACTACAATGATGCATTCAACATACGACCTATGGAGTTTGCTAATGATGTGGAGA
TGAATGCTTGGGAAACATACAGAGTGACTTCTTTGGGACCTTAACCCCAAACCTAA
CTCATGGAGAGCTTCTACAGCTCAATCTTACAATACAGCATAAGTTACTGGCTTAGAAT
ACTTAAATTTATTGAAGTTTAGTTTTCAGAGTCTACCACAAGGGTTGTTGATTCTGACGT
TATAGCAAAGAATAAAGCTCATGATTTTGGAGGGAAAGACTCTATGAGCTTGATGGGT
CCCTGAAAGGACCTCTTACAAATATTTTAAATTTTTTGTACTAGTAGAAACATAGA
TTATGAGGTGTGACTTATTATTATTTTTTACAATTGTTTGTACCTCATTGATGTATTTG
ATT

>G2134 Amino Acid Sequence (domain in AA coordinates: TBD)

MAGLRNSGNSDKAQNDGKGVPSAYRGVVRKRWKGVSEIREPGTKNRIWLGSFETPEMAA
TAYDVAAFHFRGREARLNFPELASSLPRPADSSSDSIRMAVHEATLCRTTEGTESAMQVD
SSSSSNVAPTVMRLSPREIQAINESTLGSPTMMHSTYDPMEFANDVEMNAWETYQSDFL
WDP*PQNLTHGELLQLNLTI*

>G2151 (236..1321)

TTTTTTTTTTAGGGTTCATAAGAACAAATTGGATTTTGAGCTCACAGTATAAATAACCCG
ACTTTGATTACTGGGTAATTTTAAACCGCCATTGTTGTCTCTTTACTACTTTTGGGAA
TAGGGTTTATGATTTCTGGGTATTAGATTAGATAAATTGTTTCTTTTGTGTTAATC
AATTTAAAAATCTCTATTCTGTAAAGACTTGTAATTTTGGAGTTTTAATGCATGGA
CGGAAGAGAAGCAATGGCATTTCCAGGCTCGCATTTCTCAGTACTATCTTCAAAGAGGAGC
CTTTACTAATCTCGCACCTTCCCAAGTCGCGAGTGGGCTTACGCGCCGCCGCCACATAC
GGGATTGAGGCCAATGTCTAACCTAACATTATCACCTCAGGCTAACATCCAGGACC
TCCTTTCTCGGATTTTGGACACACCATTCACATGGGAGTGGTCTCCTCTGCTTCTGATGC
TGATGTGCAACCGCCACCGCCACCGCCACCAGAGGAACCGATGGTTAAGAGGAAACG
TGGACGGCCAAGAAAGTATGGAGAACCGATGGTTAGTAATAAGTCTAGGGACTCTTCTCC
AATGTCTGATCCTAATGAACCTAAACGGGCCAGAGGTCGACCTCCTGGAAGTGAAGGAA
GCAACGCTTGCTAATCTTGGTGAGTGGATGAATACTTCAGCTGGACTTGCTTTTGCAAC
TCATGTGATCAGCATTGGAGCAGGAGAAGACATTGCTGCGAAAGTTTTGTCAATTTTACA
ACAAAGACCTCGGGCTCTTTGTATAATGTCAGGCACTGGAACCATTTCTTCAGTCACTCT
GTGCAAAACCGGTTCAACCGATCGTCACTTAACATACGAGGGACCTTTTGGATTATAAG
TTTTGGTGGATCTTATTTGGTGAATGAAGAAGGTGGATCCAGAAGTCGAACAGGCGGATT
GAGTGTCTCTCTTCTCGTCCCGATGGTAGTATTATGCCGGTGGAGTTGACATGCTTAT
CGCAGCCAACTTGTTCAAGTGGTGGCATGTAGTTTTGTATACGGAGCAAGGGCAAAGAC

TCATAATAACAATAACAAGACCATCAGACAAGAAAAGGAACCAAATGAAGAGGACAACAA
TAGTGAAATGGAGACCACACCGGGTAGTGCAGCTGAACCAGCAGCATCTGCGGGTCAGCA
GACGCCACAGAACTTCTCTTCTCAGGGAATAAGGGGGTGGCCCGTTTCAGGCTCAGGCTC
TGGCAGATCATTGACATTTGCAGAAACCCACTCACTGATTTTGATTGACTCGTGGATG
ATATACACTATTAGTCTTTGAAGCAGCAGCATACAAAATGTGATTGCTGTACATATGTTA
TTGTAGATTTCTCTCTGGGAATGTTGAAATCAGACATTTAAGGATTGATACTAGATCTCT
CAGCTCCTTCTAACATTGTTAATGTAACAGAACCTCCCACCTTTCATGCTATTTGC
>G2151 Amino Acid Sequence (domain in AA coordinates:93-113, 124-144)
MDGREAMAFPGSHSQYYLQRGAFTNLAPSQVASGLHAPPPHTGLRPMSPNIHHPQANNP
GPPFSDFGHTIHMGVVSSASDADVQPPPPPPPEEPMVKRKRGRPRKYGEPMVSNKSRDS
SPMSDPNEPKRARGRPPTGTRKQRLANLGEWMNTSAGLAFAPHVISIGAGEDIAAKVLSF
SQQRPRALCIMS GTGTISSVTLCCKPGSTDRHLTYEGPFEEISFGGSYLVNEEGSRRTG
GLSVLSRPDGSIIAGGVMDLIAANLVQVVACSFVYGARAKTHNNNNKTIHQEKEPNEED
NNSEMETTPGSAAEPASAGQQTPONFSSQIRGWPGSGSGSRSLDICRNPLTDFDLTR
G*
>G2154 (82..1317)
GCAAAAAGAAAAAATGAAAAAATCCCTAACTCTCTCTCTCTAGAAATCTTATTTTTG
TGCGTATCTCTCTAAAAAGGAATGGATCCTAACGAAAGCCACCATCACCACCAACAACAA
CAGCTCCATCACCCTCCACCAACAGCAACAGCAGCAGCAGCAACGACTCACTTCT
CCTTACTTCCACCACCAACTACAGCACCATCACCACCTTCCAACCACCGTAGCAACCACC
GCTTCTACCGGAAACGCCGTTCCATCTTCCAACAATGGGCTTTTCCCTCCGCAGCCTCAG
CCACAGCACCAGCCTAATGATGGGTCTCTCTCTCGCGGTGTACCCTCATTAGTTCCG
TCCTCGGCTGTGACGCGCGCGATGGAGCCGGTAAAGAGGAAGAGGGGTCGACCAAGAAAG
TATGTGACGCCGGAACAAGCCCTAGCGGCTAAGAAATGGCGTCTTCTGCGAGTAGTTCTG
TCTGCTAAACAGAGCGGAGAGCTTGCTGCTGTACCGGTGGTACGGTATCGACTAATTCC
GGGTCTATCCAAGAAATCTCAGCTTGGTTCTGTCTCGGAAAACTGGACAATGTTTACTCCG
CATATTGTTAATATAGCTCCTGGCGAGGATGTGGTCCAGAAAATTATGATGTTGCGAAAC
CAAAGCAAGCATGAATATGCGTTCTTCTGTCATCAGGCACTATCTAATGCATCCTTG
CGCCAACCGGCTCCATCAGGAGGCACTTACCATATGAGGGTCAATACGAGATTCTCTCA
CTATCTGGATCCTATATCCGAATGAACAAGGTGGTAAATCCGCGCGCCTTAGCGTTTCT
TTATCTGCTTCAGATGGTCAGATCATCGGTGGAGCGATTGGTAGCCATCTCACAGCTGCT
GGCCCGGTTCAGGTGATTCTTGGTACGTTTTCAGCTTGATAGAAAGAGGATGCCGCCGGG
AGTGGTGGGAAAGGGGATGCTTCAAACAGTGGAAGTCGGTTAACTTCTCCTGTAGCTCT
GGACAGTTGCTTGGCATGGGTTTCCCTCCTGGTATGGAATCTACGGGAAGAAATCCAATG
AGGGGAAACGACGAGCAACATGATCATCATCATCAAGCCGGTTTGGGTGGACCTCAT
CATTTTCATGATGCAAGCGCGCAGGGGATACACATGACACATTCCAGGCCATCTGAATGG
CGCGGAGGAGGCAACAGCGGTCTGATGGCAGAGGCGGTGGCGGGTATGATTTGTCAGGA
AGGATAGGACATGAGTCGTCGGAGAATGGAGATTACGAGCAGCAAAATACCGGATTAGCAG
AGCTTCCAGGAGAAGTGTGTAGAGTTTAGATCCCAAGTAGAGAAACAGAAGCGAGCAAA
GAATCTGAACTGAGAGAGGACTTATTAGACAGAGACTCGTCTGAAGGGTCTTTAATCATA
GAAAGAAGTTGCTGAGTGATTGCTTTTGTCTTCTTCTTGGTACGGTGTATTATATTAAAC
TCCACAACCTTTTTTTTATACTTTTCAGTAACGATTCTCCTTCACTTTCAATTTTCATTCCT
TTTTTTTATACTCTTTTTCTTTCTTATAATATTTTTTTTGGTTTTTCTTTCGTTTGTTA
CTAAAAAAGGAAATGCTCTTTTGTGAAATATATACACTTCGTTTTG
>G2154 Amino Acid Sequence (domain in AA coordinates:97-119)
MDPNESHHHHQQQLHHLHQQQQQQQQLTSPYFHHQLQHHLHLPPTTVATTASTGNAV
PSSNNGLFPPQPQPHQPNDSGLAVYPHSPVSSAVTAPMEPVKRRGRPRKYVTPEQA
LAAKKLASSASSSSAKQRRELAAVTGGTVSTNSGSSKKSQLGSVGKTGCFTPHIVNIAP
GEDVVQKIMMFANQSKHELCLVLSASGTISNASLRQPAPSGGNLPYEGQYBILSLSGSYIR
TEQGGKSGGLSVLSASDQIIGGAIGSHLTAAGPVQVILGTFQLDRKDAAGSGGKGDA
SNSGSRSLTSPVSSGQLLGMGFPFPMESTGRNPMRGNDEQHDHHHQAGLGGPHHFMMPAP
QGIHMTSRPSEWRGGGNSGHDGRGGGGYDLSGRIGHESSENGDYEQQIPD*
>G2157 (306..1238)
TCTTTTGAATTTTAACCTTTTTTTCAGTAGCAAGCCAAAAAAGAGCAAAAGAAGTT
CCTTTTATGATAAAGGTATGATGATAGCAACAAATGATACCCCATGTCTTGTGTGTCT
GCTTCATGCAACATGTTGGTTTGGATTGTTAATCTAAAAGTTAAGATAAGGTTTTTCG
GATTCTCTTCTGTCTTGTAAATAGTTTCTTGTGCGAGAGCCATCAACACCAACTTCAACA

AAAAAAACAAGAAAAAGAAAAAGATTCTCTTTCTCGTTTATTTCCATTAGAGAAGAAAA
AAAGAATGGCGAATCCTTGGTGGGTAGGGAATGTTGCGATCGGTGGAGTTGAGAGTCCAG
TGACGTCATCAGCTCCTTCTTTGCACCACAGAAACAGTAACAACAACACCCACCGACTA
TGACTCGTTCGGATCCAAGATTGGACCATGACTTCACCACCAACAACAGTGGAAGCCCTA
ATACCCAGACTCAGAGCCAAGAAGAACAAGACAGCAGAGACGAGCAACCAGCTGTTGAAC
CCGGATCCGGATCCGGGTCTACGGGTCTGTCGTCCTAGAGGTAGACCTCCTGGTTCCAAGA
ACAAACCAAGAGTCCAGTTGTTGTTACCAAAGAAAGCCCTAACTCTCTCCAGAGCCATG
TTCTTGAGATTGCTACGGGAGCTGACGTGGCGGAAAGCTTAAACGCCTTTGCTCGTAGAC
GCGGCCGGGGCGTTTCGGTGTGAGCGGTAGTGGTTTGGTTACTAATGTTACTCTGCGTC
AGCCTGCTGCATCCGGTGGAGTTGTTAGTTTACGTGGTCAGTTTGAGATCTTGTCTATGT
GTGGGGCTTTTCTTCTACGTCTGGCTCTCCTGCTGCAGCCGCTGGTTTAAACATTACT
TAGCTGGAGCTCAAGGTCAAGTTGTGGGAGGTGGAGTTGCTGGCCCGCTTATTGCCTCTG
GACCCGTTATTGTGATAGCTGCTACGTTTGAATGCCACTTATGAGAGGTTACCGATTG
AGGAAGAACAACAGCAAGAGCAGCCGCTTCAACTAGAAGATGGGAAGAAGCAGAAAGAAG
AGAATGATGATAAGAGAGTGGGAATAACGGAACGAAGGATCGATGCAGCCGCCGATGT
ATAATATGCCTCCTAATTTTATCCCAATGGTCATCAAATGGCTCAACACGACGTGTATT
GGGGTGGTCTCCGCCCTCGTCTCCTCTCGTATTGATTAGTTAGATAGGCGGTGGTTG
GTGCGTCTTTTTACTGGAATGATTATATTTTCCATTAGGATGGTTAGGCTTTTGTATT
TAAAGCTATCAAGTTTCTTTTTTTTTTACGGATAATTCGGATGACAATTAGCTAGTGT
GTTTGTGTTTGTGGCGGCTTTTCTGACTTGACTATTTTGATCGCGGATAGCTTTGTA
TGAAAGTGAATTGATTGTAGAATCGTCTTTGAATTTGATGTTGAAAAAACCA
>G2157 Amino Acid Sequence (domain in AA coordinates: 82-102, 164-107)
MANPWWVGNVAIGGVESPTSSAPSLHHRNSNNNPPTMTRSDPRLDHDFTTNSGSPNT
QTQSQEEQNSRDEQPAVEPGSGSGTGRPRGRPPGSKNPKSPVVVTKESPNSLQSHVL
BIATGADVAESLNAFARRRGRGVSVLSGSLVTNVTLRQPAASGGVSVLRGQFEILSMCG
AFLPTSGSPAAAAGLTIYLAGAQGVVGGGVAGPLIASGPVIVIAATFCNATYERLPIEE
EQQQEQPLQLEDGKKQKEENDDNESGNNGNEGSMQPPMYNMPNFI PNGHQMAQHDVYWG
GPPPRAPPSY*
>G2181 (1..1005)
ATGATGCTTGCGGTGGAAGATGTGTTAAGCGAACTCGCCGGAGAAGAAAGGAACGAGAGA
GGATTGCCACCTGGCTTCCGGTTTACCCGACGGACGAAGAGCTCATTACCTTCTACTTA
GCTTCCAAAATCTTCCATGGTGGTCTCTCCGGCATTCACTTCCGAAGTTGATCTCAAC
CGCTGTGAACCTTGGGAGCTACAGAAATGGCGAAGATGGGAGAGAGAGAGTGGTACTTT
TATAGTCTAAGGGACAGGAAATATCCGACAGGTTTGAGGACTAACAGAGCAACTACTGCT
GGTACTTGAAAGCTACCGGCAAAGATAAGGAAGTCTTCTCCGGCGGAGGAGGACAGCTT
GTTGGGATGAAGAAGACGTGTTGTTCTACAAAGGTAGGGCTCCAGTGGCCTCAAGACT
AAGTGGGTCTGTCATGATGATATCGCCTCGAAAACGACCATTCACACCGCCACAGTGTAA
GAGGAATGGGTGATTTGCAGAGTGTTCATAAAACAGGAGACAGAAAAATGTTGGATTA
ATCCATAACCAATCAGTCACTTTCATAACCATTCACTCTCAACAACACATCATCATCAT
CATGAAGCCTTACCTTTGCTTATAGAACCTTCCAACAAAACCTAACCAACTTCCCATCA
CTACTCTACGATGATCCACACAAAACCTACAATAATAACAACCTTCTTCATGGATCATCA
GGCCACAACATCGACGAGCTCAAAGCCTTAATCAACCCTGTCTCTCAGCTCAACGGT
ATCATCTTTCTTCAGGGAAACAACAACGACGAAGACGACTTCGACTTTAACCTCGGC
GTGAAAAACAGAGCAGTCTTGAACGGTAACGAAATTGACGTACGAGATTACTTGGAGAAC
CCTCTGTTTCAGGAAGCGAGTTATGGTCTGTTGGGTTTTTCGCTCTCTCTGGACCTCTT
CACATGCTACTAGATTCTCCATGTCTTTAGGATTCCAGCTGTAG
>G2181 Amino Acid Sequence (conserved domain in AA coordinates:22-169)
MMLAVEDVLSELAGEERNERGLPPGFRFHPTDEELITFYLASKIFHGGLSGIHISEVDLN
RCEPWELPEMAKMGEREWFYFSLRDRKYPTGLRNRATTAGYWKATGKDKEVFSGGGGQL
VGMKKTIVFYKGRAPRGLKTKWVMHEYRLNDSHRHTCKEEWVICRVFNKTGDRKNVGL
IHNQISYLNHNSLSTTHHHHHEALPLLI EPSNKTLTNFP SLLYDPHQNYNNNNFLHGSS
GHNIDELKALINPVVSQNLNGIIFPSGNNNNDEDDFDN LGVKTEQSSNGNEIDVRDYLEN
PLFQEASYGLLGFSSSPGLHMLLDSPCPLGFQL*
>G221 (115..795)
CTCTCTTATTCTCTCACTCTTTTTTTTTTATATTCTCTCTCTCTAAATCTATAAAATAT
ATTTAAAAAAGTATGATGATATAATAAAGTAAATAAAGAATAATAACAAAAAATGGAG
AAAAGAGGAGGAGGAAGTAGTGGAGGTTCCGGATCATCAGCAGAAGCAGAAGTGAGAAAA

GGACCATGGACGATGGAAGAAGATCTTATTCTTATCAACTATATCGCCAAACCACGGCGAT
GGTGTGTTGGAATCTCTCGCCAAATCTGCAGGTCTAAACGAACCGGGAAAAGTTGCCGG
CTCCGGTGGCTGAATATCTCCGCCCGACGTACGACGGGAAACATCACTCCAGAAGAG
CAACTTATCATCATGGAACCTCATGCTAAGTGGGGAAACAGGTGGTCGAAAATCGCCAAA
CATCTTCAGGAAGAACGGACAACGAGATCAAAAATTTCTGTAGGACAAGAATTCAAAAA
TACATCAAGCAATCGGATGTAAACAACATCGTCCGTTGGATCTCATCATAGCTCAGAG
ATCAACGATCAAGCTGCAAGCAGCTCGAGCCATAATGTCTTTGTACACAAGATCAAGCG
ATGGAGACTTATCTCTACACCGACATCATATCAACATACCAATATGGAATTCAACTAT
GGTAACTATTCGGCCGCGGACGTGACGGCAACCGTGGATTATCCAGTACCGATGACCGTT
GATGATCAAACCGGTGAAAATATTGGGGCATGGATGATATTGGTCATCAATGCATTTA
TTGAATGGTAATTGATTGATCGGTGGACAAAACATGGAATATTAATTGAGTATTATATAT
GATTTTATAGGAGTACTATTATTAGTACGTGACATGTATATGTTTTGCTCGTTGTAGAG
GTTTGGGGTTATAATTAATATATAATGTTATCTAATATGCAACCTTGATACATATTTGGA
TCTTTATTGAACCCATGTTATACATAAATAAAATTGTTGAAGGGTCATAAAAAAAAAAA
AAAAAAAAAAAAA

>G221 Amino Acid Sequence (domain in AA coordinates: 21-125)
MEKRGSSSGSSAEVRKGPWTMEEDLILINYIANHGDGVWNSLAKSAGLKRTGKS
CRLRLWNLRLPDRVRRGNITPEEQLIIMELHAKWGNRWSKIAKHLPGRTDNEIKNFCRTRI
QKYIKQSDVTTTTSSVSHSSEINDQAASTSSHNVFCTQDQAMETYSPTPTSQHTNMEF
NYGNYSAAAVTATVDYPVPMTVDDQTGENYWGMDIWSMHLNNGN*

>G2290 (119..982)

TTCTTTCTTTCTTTCTTTCTTCTTCCAATCAAGAACAAACCCCTAGCTCCTCTCTTTTCTC
TCTCTACCTCTCTTTCTCTATCTTCTCTTATCACTACTTCTCTCGCCGATCAATCATCAT
GAACGATCCTGATAATCCCGATCTGAGCAACGACGACTCTGCTTGGAGAGAACTCACACT
CACAGCTCAAGATTCTGACTTCTTCGACCGAGACACTTCCAATATCCTCTCTGACTTCGG
TTGGAACCTCCACCCTCCTCCGATCATCCTCACAGTCTCAGATTGCACTCCGATTTAAC
ACAAACCACCGGAGTCAAACCTACCACCGTCACTTCTTCTTGTTCCTCATCCGCCGCCGT
TTCCGTTGCCGTTACCTCTACTAATAATAATCCCTCAGCTACCTCAAGTTCAAGTGAAGA
TCCGCCCGAGAACTCAACCGCTCCGCCGAGAAAACACCACCACCGGAGACACCAAGTGA
GGAGAAGAAGAAGGCTCAAAAGCGAATTCCGCAACCAAGATTTCGATTTCATGACCAAGAG
TGATGTGGATAATCTGAAGATGGATATCGATGGCGTAAATATGGACAAAAGCCGTCAA
GAATAGCCCATTTCCCAAGGAGCTACTATAGATGCACAAACAGCAGATGCACGGTGAAGAA
GAGAGTAGAACGTTTCATCAGATGATCCATCGATAGTGATCACAACATACGAAGGACAACA
TTGCCATCAAACCATTTGGATTCCCTCGTGGTGAATCCTCACTGCACACGACCCACATAG
CTTCACTTCTCATCATCATCTCCCTCCTCCATTACCAAATCCTTATTATTACCAAGAACT
CCTTCATCAACTTCACAGAGACAATAATGCTCCTTCACCGCGGTTACCCCGACCTACTAC
TGAAGATACACCTGCCGTGTCTACTCCATCAGAGGAAGGCTTACTTGGTGATATTGTACC
TCAAACATATGCGCAACCCCTTGAGGTAAGCTTGGTACGTAGCAATAGCTAAGGAGGTGCTA
ACTCATTATATATAGAAGATATTGCAGACGAGAATATGCGCAGGGAGGGGTATAACAATAT
GGCGTTGTAACAATGGATCTATATATTACCTCATTTGTTGATCAATAGCACACCACCGGTA
CGTTTGCAATTTCTTCATGTATATTCTTGTATATATGTAGTTATATATCCAGGTATAA
TTTTGATGTAACACAACATTAATCTTAATCGTGGATCCATCCACATTTGATGCATGTAT
GTGCACTTAAGAAAAAGAACATGGAGGAAATAACGTATTTTTTTTATTATTCT

>G2290 Amino Acid Sequence (conserved domain in AA coordinates:147-205)

MNDPDNPDLSNDDSAWRELTLTAQDSDFDRDTSNLSDFGWNLHHSSDHPHSLRFDSDL
TQTTGVKPTVTSSCSSSAVSVAVTSTNNNPSATSSSSSDPAENSTASAECTPPPETPV
KEKKKAQKRIRQPRFAFMKSDVDNLEDGYRWRKYQKAVKNSPFPRSYRCTNSRCTVK
KRVERSSDDPSIVFTTYEQHCHQTIGFPRGGILTAHDPHSFTSHHHLPPLPNPYYYQE
LLHQLHRDNNAPSRLPRPTTEDTPAVSTPSEEGLLGDIVPQTMRNP*

>G2299 (231..941)

GCCAAAATTTTACCAACATTTTCTCTTCTCATATCAAAGTTTCTCTCTCATTTCTTCAT
CACACTTCACTGCCCTGTTTTTTTTCTCATTTTGAATAGTTCTCAAACCTATATATTTT
TCCCCCTGAAGCCTAGCTATTTCTTTTATTGTCATTAATCTCGGGATCCGAATCGAAAA
AAGCAATCAGAATACTGTACGATACTTGTGCCTAAGCTAACACAATGGCAGAGG
AATACTACAGCCTCCGCTCGGAGAGAGTAACTCAGCTTCTTGTCCCTAACTCGGAGTCTG
ACTCAGTGAGTGACAAAAGCAAAGCTGAGCAAAGCGAGAAGAAGACTAAACGTGGGAGAG
ACTCCGGTAACACCCTGTTTATCGCGGAGTAAGGATGAGGAAGTGGGGAAAATGGGTGT

CGGAGATTTCGTGAGCCGAGGAAGAAATCACGTATTTGGCTGGGAACCTTTCCCGACGCCGG
AGATGGCGCGCGCTGCACACGACGTGGCGGCTCTGAGCATTAAAGGAACGGCCGCTATAC
TAAACTTCCCTGAACTCGCTGACTCATTCCCTCGACCCGTTTCATTAAGCCCTCGAGACA
TTCAGACAGCAGCTCTTAAAGCAGCTCACATGGAACCGACGACGTCGTTTTTCATCTTCCA
CGTCTTCGTCGTCGCTTTTGTCTTCTACGTCTTCGCTCGAGTCTCTTGTGTTGGTGATGG
ACCTCTCGAGGACTGAGTCGGAGGAGCTCGGTGAGATTGTGGAGCTTCCAAGTCTCGGGG
CGAGTTACGACGTCGACTCGGCTAACCTTGGGAACGAGTTTGTCTTCTATGACTCAGTTG
ACTACTGTTTTATATCCGCCGCCGTGGGGACAGTCGTCCGAAGATAACTATGGTCACGGAA
TTAGCCCTAATTTTGGCCATGGCTTGTCTATGGGATCTCTAACAGTTTATTTTGTATCATT
ACCATAATGTTTTGTTTAAACAGTTTATTTTGTATCATTGCCATAATGTTTTGTTTAAAT
CACGTTTTTAAACCCCTTTGCTGTTTTTGTTTTTTTTTTGAGTTTTT
>G2299 Amino Acid Sequence (conserved domain in AA coordinates:48-115)
MAEYYSLRSERVQLLPNSESDSVSDKSKAEQSEKKTKRGRDSGKHPVYRGVVRMRNWG
KWVSEIREPRKKSRIWLGTFPTPEMAARAHDAALSIKGTAAILNFPPELADSFPRPVSL
PRDIQTAALKAAHMEPTTSFSSSTSSSSSLSTSSLESVLVMDLSRTESEELGEIVELP
SLGASYDVDSANLGNFVFDYDSVDYCLYPPPWGQSSSEDNYGHGISPNFGHGLSWDL*
>G2340 (274..1275)
ATACAAACTCCCTCTTCTCTATCTTCTTCATCTTAAAGAAAAAATAAGAGATATTCGTA
AAGAGAGAACAACAAATTTTCACTTACGAAAAGCTAGCAAAGTCGAGTATCGAGGAATAA
CAGAATAAGACGTATCTATCCTTGCCTTAATGTTCTTACCAAAGATCTAGTCCTTTCTT
TGATGATCGATCCATCACAAAGCCACAACAACAACACTACATCTCTTCTCTATCTCT
AGCTTCTATTTTAAATACATTCAAGAATCAAGAATGGTACGGACGCCGTGTTGTAGAGCA
GAAGGGTTGAAGAAAGGAGCATGGACTCAAGAAGAAGACCAAAGCTTATCGCCTATGTT
CAACGACATGGTGAAGGCGGTGGCGAACCCCTCCGGACAAAGCTGGACTCAAAAGATGT
GGCAAAGCTGCAGATTGAGATGGGCGAATTACTTAAAGACCTGACATTAAACGTGGAGAG
TTAGCCAAGACGAGGAAGATTCCATCATCAACCTCCACGCCATTATGGCAACAAATGG
TCGGCCATAGCTCGTAAATAACCAAGAAGACAGACAATGAGATCAAGAACCATTGGAAC
ACTCACATCAAGAAATGTCTGGTCAAGAAAGGTATTGATCCGTTGACCCACAAATCCCTT
CTCGATGGAGCCGGTAAATCATCTGACCATCCGCGCATCCCGAGAAAAGCAGCGTTCAT
GACGACAAAGATGATCAGAATTCAAATAACAAAAAGTTGTCAGGATCATCATCAGCTCGG
TTTTTGAACAGAGTAGCAAAACAGATTCCGGTCATAGAATCAACCACAATGTTCTGTCTGAT
ATTATTGGAAGTAATGGCCTACTTACTAGTCACACTACTCCAACACAAAGTGTTCAGAA
GGTGAGAGGTCAACGAGTTCTTCTCCACACATACCTCTTCAATCTCCCATCAACCGT
AGCATAACCGTTGATGCAACATCTCTATCCTCATCCAGTTCTCTGACTCCCCCGACCCG
TGTTTATACGAGGAAATAGTCGGTGACATTGAAGATATGACGAGATTTTCATCAAGATGT
TTGAGTCATGTTTTATCTCATGAAGATTTATTTGATGTCGGTTGAGTCTGTTTGGAGAAT
ACTTCATTGATGAGGGAATTAATGATGATCTTTCAAGAGGATAAAATCGAGACGACGTCG
TTAATGATAGCTACGTGACGCCGATCAATGAAGTTGATGACTCCTGTGAAGGGATTGAC
AATTATTTTGGATGAGTTATATTGATGATGATGAAAATTTGCATTTGGCATGTAAATCAA
TTAGAGTTTGATTTGCTATGGTGTGTTTTAGTTTGTGTGTGTAGTGTGTTTCGACCGTCAA
AAAAAAAAAAAAAAAAAAAAAAAAAAAA
>G2340 Amino Acid Sequence (domain in AA coordinates:14-120)
MVRTPCCRAEGLKKGAWTQEEDQKLIAYVQRHGEQWRTLPDKAGLKRCKGKSCRLRWANY
LRPDIKRGEFSQDEEDSIINLHAIHGNKWSAIARKIPRRIDNEIKNHNTHIKKCLVKKG
IDPLTHKSLLDGAGKSSDHSAPHEKSSVHDDKDDQNSNNKLSGSSSARFLNRVANRFGH
RINHNVLSDIIGSNGLTSHHTPTTSVSEGERSTSSSSTHTSSNLPINRSITVDATSLSS
STFSDSPDPCLYEEIVGDIEDMTRFSSRCLSHVLSHEDLLMSVESCLENTSFMREITMIF
QEDKIETTSFNDSYVTPINEVDDSCGIDNYFG*
>G2346 (1..1011)
ATGGAGTTGTTAATGTGTTCCGGGTGAGCCGAGTCAGGTGGTTCTTCTTCCACCGAGTCT
TCTTCACTCAGTGGTGGACTCAGGTTTGGTCAGAAGATCTACTTCGAGGATGGATCCGGA
TCCAGAAGCAAGAACCGGTCATACCGTTTCGTAAGTCGTCTACCACGGCGAGGTGCCAA
GTGGAAGGTTGTAGAAATGGATCTAAGCAATGTTAAAGCTTATTACTCGAGACACAAAGTT
TGTTGCATTCACTCTAAATCATCTAAAGTCATTGTCTCTGGTCTTCATCAAAGGTTTTGT
CAACAATGTAGCAGGTTTACCAGCTTCTGAGTTTGACTTGGAGAAAAGAAGTTGTCGC
AGAAGACTCGCTTGTCAACGAACGACGAAGAAAACCAACCCACAACCGGCTCTTTTC
ACTTCTCATTACTCTGAATCGCTCCATCTCTTTACGGAACCCCAATGCTGCAATGATT

AAAAGCGTTTTGGGAGATCCTACTGCGTGGTCAACCGCAAGATCAGTGATGCAGCGGCCT
GGACCGTGGCAGATTAATCCAGTTAGGGAAACCCATCCACACATGAATGTTTTATCACAT
GGAAGCTCAAGCTTTACTACATGTCCAGAGATGATAAACACAATAGCACAGATTCAAGC
TGTGCTCTCTCTCTCTCTGTCAAACTCATACCCAATTCATCAGCAGCAACTTCAGACACCA
ACAAATACATGGCGACCATCTTCTGGTTTCGACTCGATGATCTCATTCTCCGATAAGGTT
ACAATGGCTCAGCCACCGCCCATTTCAACCCATCAGCCGCCCATCTCAACACATCAGCAG
TACCTCAGCCAAACTTGGGAAGTCATCGCGGGCGAAAAGAGCAATTCACATTATATGTCT
CCTGTGAGTCAAATCTCGGAGCCAGCAGATTTCAGATAAGCAATGGCAGTGTGTCGCC
TATTCTCCTCCGTCCTTACTATCTCTTGTGTGCTACTTGCGCCGCTATAG

>G2346 Amino Acid Sequence (domain in AA coordinates: 59-135)
MELLMCSGGAESGSSSTESSLSGGLRFGQKIYFEDGSGSRSKNRVNTVRKSSTTARCO
VEGCRMDLSNVKAYYSRHKVCCIHSKSSKVI VSGLHQRFCQQCSRHFHQLSEFDLEKRSCR
RRLACHNERRRRKPQPTTALFTSHYSRIAPSLYGNPNAAMIKSVLGDPTAWSTARSVMQRP
GPWQINPVRETHPHMNVLSHGSSSFTTCPEMINNNSTDSSCALSLSNSYPHQQQLQTP
TNTWRPSSGFDMSISFSDKVTMAQPPPISHTQPPISHTQQYLSQTWEVIAGEKSNSHYMS
PVSQISEPADFQISNGSVSPYSPPSLLSLVCYLRLP*

>G237 (1..852)
ATGGCGAAGACGAAATATGGAGAGAGACATAGGAAAGGGTTATGGTCACCTGAAGAAGAC
GAGAAGCTAAGGAGCTTCACTCTCTTATGGCCATTCTTGCTGGACCATGTTCCCATC
AAAGCTGGGTTACAAAGGAATGGGAAGAGCTGCAGATTAGATGGATTAAATTACCTAAGA
CCAGGGTTAAAGAGGGATATGATTAGTGCAGAAGAAGAAGAGACTATCTTGACGTTTCAT
TCTCCCTTGGGTAACAAGTGGTGCAGAAATAGCTAAATCTTACCGGGAAGAACAGACAAT
GAGATAAAGAACTATTGGCACTCTCATTGAAAAAGAAATGGCTCAAGTCTCAGAGCTTA
CAAGATGCAAAATCTATTTCCTCTCTCGTCTTCATCATCATCATTGTTGCTTGTGGA
GAAAGAAATCCGGAACCTTGATCTCGAATCACGTGTTCTCCCTCCAGAGACTTCTAGAG
AACAAATCTTCATCTCCCTCACAAGAAAGCAACGGAATAACAGCCATCAATGTTCTTCT
GCTCCTGAGATTCCAAGGCTTTTCTCTCTGAATGGCTTTCTTCTCATATCCCCACACC
GATTATTCTCTGAGTTTACCGACTCTAAGCACAGTCAAGCTCCAAATGTGGAAGAGACT
CTCTCAGCTTATGAAGAAATGGGTGATGTTGATCAGTTCATTACAACGAAATGATGATC
AACAAAGCAACTGGACTCTTAACGACATTGTGTTTGGTTCCAAATGTAAGAAGCAGGAG
CATCATATTTATAGAGAGGCTTCAGATTGTAATCTTCTGCTGAATCTTTTCTCCACCA
ACAACGACGTAAATTGCGTTTATTGTAATGTAAATCAAATTTCTAAGGCAAAACCGGAAA
AAAAAAAAAAAAAAAAAAAAA

>G237 Amino Acid Sequence (domain in AA coordinates: 11-113)
MAKTKYGERHRKGLWSPEDEKLRSFILSYGHSCWTTVPIKAGLQRNGKSCRLRWINYLR
PGLKRDMSIAEEETILTFSPLGNKWSQIAKFLPGRDNEIKNYWHSLSLKKKWLKSQSL
QDAKSISPPSSSSSSSLVACGERNPETLISNHVFSLQRLLENKSSSPSQESNGNNSHQCSS
APEIPRLFFSEWLSSSYPHTDYSSEFTDSKHSQAPNVEETLSAYEEMGDVDQPHYNEMMI
NNSNWTLNDIVFGSKCKQEHIIYREASDCNSSAEFFSPPTT*

>G2373 (48..1199)
GCAAAATCCTCAGATCGTCTTACCTTCTCCGAATCGATCGATTTTTTCATGGAGGACGACG
ACGAGATTCACTCAATCCATCTCCGGGAGATTCTTCCCTTTACCACAAGCTCCTCCTT
CTCCGCCGATTGTTGCCAACAAACGACGTGACGGTGGCCGTCGTGAAGAAACCACAACCGG
GGCTTTCTTCTCAATCTCCGTCCATGAACGCTTTAGCGTTAGTGGTTCACTCCTTCTG
TAACCGGTGGTGGTGGTAGCGGAAACAGAAACGACGAGGAGGAGGAGGAAGCGGTG
GTGGTGGAGGAGGAAGAGATGATTGTTGGAGCGAAGAAGCTACAAAGGTTCTAATCGAAG
CTTGGGGAGATCGATTCTCTGAACAGGTAAAGGAACCTTGAAGCAACAACATTGGAAAG
AAGTAGCTGAGATTGTGAACAAGAGTCGTCAATGCAAAATACCCTAAAATGATATTCACT
GTAAGAACAGAAATTGATACGGTGAAGAAGAAGTATAAGCAAGAGAAAGCTAAGATTGCTT
CTGGTGATGGACCTAGTAAATGGGTTTTCTTCAAGAAGCTTGAGAGTTGATTGGTGGTA
CTACAACATTCACTGCTTCTTCAAAAGCTTCAGAGAAGGCTCCTATGGGAGGAGCTCTTG
GGAATAGCCGTTTCGAGTATGTTTAAACGGCAAACTAAAGGTAATCAGATTGTGCAGCAAC
ACAAGAGAAGAGAGGCTCTGATTGATGCGGTGGCATTCTTAGGAAACGTAGTGCTTCTG
AGACTGAGTCTGAGTCTGATCCTGAACCTGAGGCTTCTCTGAGGAATCTGCTGAGAGTC
TCCACCTTTGCAACCGATTCAACCGCTTTCGTTTTCATATGCCAAAGCGGTTGAAGGTGG
ATAAGAGTGGAGGTGGAGGAGTGGAGTGGAGATGTGGCGAGGGCGATACTTGGATTTA
CGGAAGCTTATGAGAAGGCGGAAACTGCTAAGCTTAAGTTAATGGCGGAACCTGGAAAAGG

AGAGGATGAAATTTGCTAAAGAGATGGAGTTGCAGAGAATGCAGTTCTTGAAAACCTCAAT
TGGAGATAACACAGAACATCAAGAAGAGGAAGAGAGGAGCAGGCAGCGAGGAGAAAGGA
GGATCGTTGATGATGATGATGATGATCGCAATGGCAAGAATAACGGCAATGTAAGTAGCTGAC
AATTGAACACACAAATGTTCTATGATATTTGCTATGATAAGCTGGATTTTAGGTTTTGA
TGG

>G2373 Amino Acid Sequence (domain in AA coordinates:290-350)
MEDDDEIQSIPSPGSSLSPPQAPPSPILPTNDVTVAVVKKPQPLSSQSPSMNALALVV
HTPSVTGGGSGNRNGRGGGGSGGGGGGRDDCWSEETKVLIWAGDRFSEPGKGLTKQ
QHWKEVAEIVNKSQCKYKPKTDIQCKNRIDTVKKKYKQEKAKIASGDGSPKWWFFKKLES
LIGGTTTTFIASSKASEKAPMGGALGNSRSMFKRQTKGNQIVQQQEQKRGSDSMRWHFRK
RSASETESESDPEPEASPEESAESLPPLQPIQPLSFHMPKRLKVDKSGGGGSGVGDVARA
ILGFTAEYEAETAKLKLMAELEKERMKFAKEMELQRMQFLKTQLEITQNNQEEERSRQ
RGERRIVDDDDDRNGKNNGNVSS*

>G2376 (39..1370)

CACGAGCTTCTGACTCAGATCCGGCGATATCGAATTCATGGAGGACGATGAAGACATCC
GATCTCAGGGTTCCGATTACCTGATCCGTCTTCTCCCCCGCGGGACGAATCACGG
TTACGGTGGCTTCGGCAGGTCCGCCTTCTTATTCTCTGACTCCTCCGGGTAAATTCGTGCG
AGAAGGATCCGGATGCGTTGGCTCTGGCGCTGCTTCCGATTACAGCCAGCGGTGGAGGGA
ATAACAGCAGTGGGAGACCAACCGGCGGGCGGGAGGGAGGATTGTTGGAGCGAAGCAG
CTACGGCTGTGTTGATTGATGCGTGGGGTGAGAGATACTTGGAGCTTAGCAGAGGGAATC
TGAAGCAGAAGCACTGGAAGAGGTGGCTGAGATTGTGAGCAGCAGAGAGGATTACGGTA
AAATTCCCAAACTGATATACAGTGTAAGAAATAGGATCGATACGGTGAAGAAGAAGTATA
AACAAGAGAAGGTGAGAATCGCTAACGGCGGTGGCCGTAGCAGATGGGTGTTCTTCGACA
AGCTTGACCGTCTGATTGGATCAACGGCGAAGATCCCGACGGCAACTTCTGGAGTCAGCG
GTCCTGTGCGGAGGATTGCATAAGATTCTATGGGTATTCCAATGGGAAGTCGTTTCAATC
TGTACCATCAGCAAGCTAAGGCTGCAACACCGCCTTTCAATAATCTTGACCGGTTAATTG
GAGCTACGGCTAGAGTCTCAGCTGCTTCTTTCGGTGGCAGTGGTGGAGGAGGCGGAGGAG
GATCTGTCAATGTACCTATGGGAATTCGATGAGTAGCCGTTTCACTCCGTTTGGACAGC
AAGGAGGACTCTGCCACAGCAAGGTAGGACACTGCCACAGCAACAGCAGCAAGGGATGA
TGGTGAAAAGGTGTAGTGAGTTCGAAACGCTGGCGTTTCAGGAAGAGGAACGCTTCTGATT
CAGACTCGGAATCTGAAGCAGCAATGTGATGATTCCGGTGACAGTTTACCACCTCCTC
CTCTGTGCAAGAGGATGAAGACGGAGGAGAAGAAGAAGCAAGATGGTGATGGAGTGGGA
ACAAATGGAGGGAGCTGACTCGGGCAATCATGAGATTCCGTGAAGCTTATGAGCAACAG
AGAATGCGAACTGCAACAGGTGGTTGAGATGGAGAAAGAGAGGATGAAGTTCTTGAAGG
AGCTTGAGTTGCAGAGAATGCAGTCTTTGTGAAGACTCAATTGGAGATATCACAACTTA
AGCAGCAACATGGGAGGAGAATGGGAAACACCAGTAATGATCATCATCAGCCGCAAGA
ACAACATCAATGCGATTGTGCAACAACAACGATTGGGTAATAACTAGAATTTAGTGA
TGCAGTGTGCTAATTGATATATTTTAGATTTGAG

>G2376 Amino Acid Sequence (domain in AA coordinates:79-178, 336-408)

MEDDEDIRSQSDSPDPSSPPAGRITVTVASAGPPSYSLTPPGNSSQKDPDALALALLP
IQASGGGNNSSGRPTGGGREDCWSEATAVLIDAWGERYLELSRGNLQKHWKEVAEIV
SSREDYKIPKTDIQCKNRIDTVKKKYKQEKVRIANGGGRSRWVFFDKLDRDIGSTAKIP
TATSGVSGPVGGLHKIPMGIPMGSRSNLYHQAKAATPPFNNLDRDIGATARVSAASF
SGGGGGGGSVNVPMPGIPMSRSAPFGQQGRTLPPQQGRTLPPQQQQMMVKRCSESKRWF
RKRNASDSDSESEAMSDSDSLPPPLSKRMKTEKKKQDGDGVGNKWRELTRAIMRF
GEAYEQTENAKLQQVVEKERMKFLKELELQRMQFFVKTQLEISQLKQGHRRMNTSN
DHHHSRKNINAIIVNNNDLGN*

>G24 (194..724)

CGGACGGTGGGCAAATATTTAAATAAAAAGTGTCCGGTGAATTCTCAATCTTTGTCTTCT
TTCGTCTCTCTTTAAACTCCTCCGTCCCTCCTTATTATGTAACCGTCTCGCCGTCAA
TTTTCAAATCTCTCCCTCCGTTCAAAACCCAGATCGAAATTTATGGTTTTGTAATTTT
TTTACCGCGCGTTATGGAGACGGAAGCGGCGGTGACAGCGACGGTTACGGCGGACGAT
GGGGATTGGGACGAGGAAGAGAGATCTGAAACCGTATAAAGGAATACGAATGAGGAAATG
GGGGAATGGGTGGCGGAGATACGGGAACCGAATAAGAGATCAAGGATCTGGTTAGGTTT
TTATGCGACGCTGAAGCGGCGGCGAGAGCTTACGACACTGCTGTTTTTACCTCCGTGG
TCCTTCAGCGAGGCTTAATTTTCCGGAGCTTTTGGCTGGACTTACTGTTTCTAACGGCGG
AGGAAGAGGTGGTGATTATCGGCGCGGTATATTAGGAGAAAAGCGGCGGAGGTTGGTGC

TCAGGTTGATGCGCTTGGAGCGACGGTGGTTGTGAATACCGGCGGCGAGAATCGCGGTGA
TTACGAGAAGATTGAGAATTGTCGTAAGAGCGGTAACGGGTCATTGGAACGGGTCGATTT
GAATAAATTACCCGACCCGAAAATTTCGGATGGTGATGATGACGAATGTGTGAAAAGAAG
ATAGAAAAAATAAAAAGTAGTTGTAGAAGGAGAGACGAGAATGTTTGTCTTTAAGATGCG
CTGTTGCCGCTAACATGCGCTTTTCGATTTTAGTGTTAAACATGCGCCTCCATTGTTTTTG
GGTTTTGTTTTCGTCGTCGATAATCAAAGATTTTAAACACAATTCTCAAATTTTTCCT
TGTACAAACTAGATTTGCATGATCTTTGTATTAACGAATAACGATTAAGTCCTAAA

>G24 Amino Acid Sequence (domain in AA coordinates: 25-93)

METEAAVTATVTAATMGIGTRKRDLPYKGI RMRKWKWVAEIREPNKRSRIWLGSYATP
EAAARAYDTAVFYLRGPSARLNFPELLAGLTVSNGGGRGDLAAYIRRKAAEVGAQVDA
LGATVVVNTGGENRGDYEKIENCRKSGNSLERVDLNLKLPDPENSDGDDDECVKRR*

>G2424 (1..999)

ATGAGGATGGAGATGGTGCATGCTGACGTGGCGTCTCTCTCCATAACACCTTGCTTCCCCG
TCTTCTTTGTCTTCGTCCTCACAATCATCTATAACCAACAACAACATTGTATCATGTGCG
GAAGATCAACACCATTTCGATGGATCAGACCACTTCATCGGACTACTTCTCTTTAAATATC
GACAATGCTCAACATCTCCGTAGCTACTACACAAGTCATAGAGAAGAAGACATGAACCCCT
AATCTAAGTGATTACAGTAATTGCAACAAGAAAGACACAACAGTCTATAGAAGCTGTGGA
CACTCGTCAAAAGCTTCGGTGTCTAGAGGACATTGGAGACCAGCTGAAGATACTAAGCTC
AAAGAACTAGTCGCGCTCTACGGTCCACAAAACCTGGAACCTCATAGCTGAGAAGCTCCAA
GGAAGATCCGGGAAAAGCTGTAGGCTTCGATGGTTTAAACCACTAGACCCAAGGATAAAT
AGAAGAGCCTTCACTGAGGAAGAAGAAGAGAGGCTAATGCAAGCTCATAGGCTTTATGGT
AACAAATGGGCGATGATAGCGAGGCTTTTCCCTGGTAGGACTGATAATTCTGTGAAGAAC
CATTGGCATGTTATAATGGCTCGCAAGTTTAGGGAACAATCTTCTTCTTACCGTAGGAGG
AAGACGATGGTTTCTCTTAAGCCACTCATTAAACCTAATCCTCACATTTTCAATGATTTT
GACCTTACCCGCTTAGCTTTGACCCACCTTGCTAGTAGTGACCATAAGCAGCTTATGTTA
CCAGTTCTTGTCTTCCAGGTTATGATCATGAAAATGAGAGTCCATTAAATGGTGGATATG
TTCGAAACCCAAATGATGGTTGGCGATTACATTGCATGGACACAAGAGGCAACTACATTC
GATTTCTTAAACCAACCGGGAAGAGTGAGATATTTGAAAGAATCAATGAGGAGAAGAAA
CCACCATTTTTCGATTTTCTTGGGTTGGGGACGGTGTGA

>G2424 Amino Acid Sequence (conserved domain in AA coordinates:107-219)

MRMEMVHADVASLSITPCFPSSLSSSSHHYNNQQQHCHMSDQHHMSDQTTSSDYFSLNI
DNAQHLRSYTTSHREEDMNPNSLSDYSNCKKDTTVYRSCGHSSKASVSRGHWRAEDTKL
KELVAVYGPQNWNLIAEKLQGRSGKSCRLRWFNQLDPRINRRAFTEEEERLMQAHRLYG
NKMAMIARLPFGRDTSVKNHWHVIMARKFREQSSSYRRRKTMSVSLKPLINPNPHIFNDF
DPTRLALTHLASSDHKQLMLPVPCFPGYDHENESPLMVDMFETQMMVGDYIAWTQEATTF
DFLNQTKSEIFERINEEKKPPFFDFLGLGTV*

>G2505 (1..1026)

ATGGGTTCTTCGTCGAACGGAGGAGTGCCACCTGGTTTCCGGTTTTCATCCGACGGACGAA
GAGCTTCTCCATTACTACTTGAAGAAGAAAATCTCTTACCAAAAGTTTGAGATGGAAGTC
ATCAGAGAGGTTGACTTAAACAAGCTTGAGCCTTGGGATTGCAAGAGAGATGCAAGATA
GGATCAACACCACA AAACGAATGGTACTTCTTCAGCCACAAGGACAGGAAATATCCGACG
GGGTCAAGGACCAACCGTGCTACTCATGCAGGTTCTGGAAGGCGACGGGACGTGACAAG
TGCAATAAGGAACCTTACAAAAGATAGGAATGAGGAAGACACTTGTGTTCTACAAAGGT
AGAGCTCCTCATGGCCAAAAGACTGATTGGATCATGCATGAGTACCGTCTTGAAGACGCT
GATGATCCTCAAGCCAACCTAGTGAAGATGGATGGGTGGTATGTAGAGTGTTTATGAAG
AAAAATTTGTTCAGGTAGTAAATGAAGGTAGCTCAAGCATTAACCTCATTGGACCAACAC
AACCATGACGCATCTAACAACAACCATGCACCTCAAGCTCGTAGCTTTATGCACCGAGAC
AGTCCATACCAGCTAGTACGTAACACGGAGCCATGACATTGCAACTTAACAAGCCTGAC
CTTGCTCTTCATCAATACCCACCAATCTTCCACAAGCCACCTTCATTGGATTTGACTAC
TCTTCAGGACTTGCAAGGGACAGTGAGAGTGCGGCTAGTGAAGGGTTACAATACCAGCAA
GCGTGTGAGCCGGGTTTAGACGTTGGTACATGTGAGACAGTGGCTAGTCATAATCATCAA
CAAGGCTTAGGTGAATGGGCAATGATGGATAGACTTGTGACTTGTACATGGGAAATGAA
GATTCCTCTAGAGGGGATTACGTATGAGGATGGTAACAACAATTGCTCCTCTGTGGTTTCA
CCAGTTCCCGCGACGAACAGCTAACATTGCGTAGTGAGATGGATTTCTGGGGTTATTCT
AAATAG

>G2505 Amino Acid Sequence (domain in AA coordinates: 10-159)

MGSSSNGGVPPGFRFHPTDELLHYLKKKISYQKFEMEVI REVDLNLKLEPWLQERCKI

GSTPQNEWYFFSHKDRKYPTGSRTNRATHAGFWKATGRDKCIRNSYKKIGMRKTLVIFYKG
RAPHGQKTDWIMHEYRLDADDPQANPSEDCGWVVCVFMKKNLFKVVNEGSSSINSLDQH
NHDASNNNHALQARSFMHRDSPYQLVRNHGAMTFELNKPDLALHQYPPIFHKPPSLGFDY
SSGLARDSESAASEGLQYQACEPLDVGTCTVASHNHQOGLGEWAMMDRLVTCMHGNE
DSSRGITYEDGNNSSSSVVQPVVPATNQLTLRSEMDFWGYSK*

>G2512 (64..798)

AACCTAGTGCCACTTAGACACAATAAGAAAACCGTTAACAAGAAGAAAAAAGATCG
AAAATGGAATATCAAACCTAACTTCTTAAGTGGAGAGTTTCCCGGAGAACTCTTCTTCA
AGCTCATGGAGCTCACAAGAATCATCTTGTGGGAAGAGAGTTTCTTACATCAATCATTT
GACCAATCCTTCTTTTATCTAGCCCTACTGATAACTACTGTGATGACTTCTTTGCATTT
GAATCATCAATCATAAAGAAGAAGAAAAGAAGCCACCGTGGCGGCCGAGGAGGAGGAG
AAGTCATACAGAGGAGTGAGGAAACGGCCGTGGGGGAAATTTCGGCGCCGAGATAAGAGAC
TCAACGAGGAAAGGGATAAGAGTGTGGCTTGGGACATTCGACACCGCGGAGGCGCGGCT
CTCGCTTATGATCAGGCGGCTTTCGCTTTGAAAGGCAGCCTCGCAGTACTCAATTTCCCC
GCGGATGTCGTTGAAGAATCTCTCCGAAGATGGAGAATGTGAATCTCAATGATGGAGAG
TCTCCGGTGATAGCCTTGAAGAGAAAACACTCCATGAGAAACCGTCTAGAGGAAAGAAG
AAATCTTCTTCTTCTTCGACGTTGACATCTTCTCCTTCTTCTCCTCCTCCTATTCATCT
TCTTCGTCTTCTTCTTCTTGTCTGCAAGAAGTAGAAAACAGAGTGTTGTTATGACGCAA
GAAAGTAATACAACACTTGTGGTCTTGTAGGATTTAGGTGCTGAATACTTAGAAGAGCTT
ATGAGATCATGTTCTTGATAATCTCTGCTTCTACATTTTTATGTAATTTGA

>G2512 Amino Acid Sequence (conserved domain in AA coordinates:79-139)

MEYQTNFLSGEFSPENSSSSWSSQESFLHQSFDQSFLSSPTDNYCDDFFAFEF
SSIIKEEGKEATVAEEEEESYRGVVRKRPWGKFAAEIRDSTRKGIRVWLGTFTAEAAAL
AYDQAAFALKGSLAVLNFPADVVEESLRKMENVNLNDGESPVIALKRKHSMRNRPGRKKK
SSSSSTLTSSPSSSSSYSSSSSSSSLSRSRKQSVVMTQESNTTLVVLEDLGAELYLEELM
RSCS*

>G2513 (69..698)

TTTCAACAGTAATTTAAGTTAACCGGAGTCTCTTTTTGTTTTCCGGCGAATTTTTGGTAC
TTTGAGTTATGAATAATGATGATATTCTGGCGGAGATGAGGCCTAAGAAGCGTGCGG
GAAGGAGAGTGTTTAAGGAGACACGTACCCAGTTTACAGAGGCATAAGGCGGAGGAACG
GTGACAAATGGGTCTGCGAAGTCAGAGAACCGACGCACCAACGCCGCATTGGCTCGGGA
CTTATCCACAGCAGATATGGCAGCGCGTGCACACGACGTGGCGGTTTGTAGCTCTGCGTG
GGAGATCCGCATGTTTGAATTTCCGCCGACTCCGCTTGGCGGCTTCCGGTGCCGGAATCCA
ATGATCCGGATGTGATAAGAAGAGTTGCGGCGGAAGCTGCGGAGATGTTTAGGCCGGTGG
ATTTAGAAAGTGGAATTACGGTTTTGCCTTGTGCGGGAGATGATGTGGATTTGGGTTTTG
GTTCCGGTTCCGGCTCTGGTTCCGGATCCGAGGAGAGGAATCTTCTTCGTATGGATTTG
GAGACTACGAAGAAGTCTCAACGACGATGATGAGACTCGCGGAGGGGCCACTAATGTCCG
CGCGCGATCGTATATGGAAGACATGACTCCTACTAATGTTTACCGGAAGAAGAGATGT
GTTATGAAGATATGTATTGTGGAGTTACAGATATTAAGTGGGACTCACATATCTACTAT
ACATAATATTTAGCTTTTATGTAAGAGGTATTTATGTGAGTTTTAAGATTGTAGATGTGT
CCCAGCGTTAGAAGTTTCTTGATGGTATGGAATCTTTGTACCTATAAAATTATAAAAT
T

>G2513 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNDIILAEMRPKKRAGRRVFKETRHVPYRGIRRRNGDKWVCEVREPTHQRRRIWLGTYP
TADMAARAHDAVLALRGRSACLNFAWSAWRLPVPESNDPDIRRVAAEAAEMFRPVDLE
SGITVLPCAGDDVDLFGSGSGSGSGSEERNSSSYGFGDYEEVSTTMMRLAEGPLMSPPR
SYMEDMTPNTVYTEEEMCYEDMSLWSYRY*

>G2519 (83..691)

CAAAGTGAAGACATAAGATCATCTTCTCGTTGATAGATCAATATAGGAACTCCAGAAGA
GAATCTTGATCAATTAAGTATCATGTCTCACATCGTGTGAAAGGAATCGAAGAAGGCA
AATGAACGAGCATCTTAAATCCCTTCGTTCTTTGACTCCTTGTCTTACATCAAAAGGGG
AGATCAAGCTTCGATCATCGGAGGAGTGATAGAGTTTCATCAAAGAGTTGCAGCAATTGGT
TCAAGTTCTTGAGTCCAGAAACGTCGAAAGACCCATAAACCAGCATCTTCCCTTATGA
TCACCAGACAATCGAGCCATCCAGTTTAGGAGCCGCCACTACCCGAGTACCGTTTAGTCG
AATCGAAAATGTGATGACCACAAGTACTTTCAAGGAAGTAGGAGCATGCTGTAACCTCCCC
TCATGCTAACGTAGAAGCAAGATTTCAGGTTCTAATGTTGTATTGAGAGTTGTCTCTAG
GCGAATCGTGGGGCAGCTCGTAAAGATCATCTCTGTCTTAGAGAAGCTATCTTTTCAAGT

TCTTCACCTCAATATTAGTAGCATGGAGGAGACTGTCTTATACTTTTTCGTTGTTAAGAT
AGGATTGGAGTGTCACTTAAGCTTGGAGGAGCTAACTCTGAAGTTCAGAAAAGCTTTGT
GTCTGATGAAGTGATCGTCTCTACCAATTAAAAACAAAATTCTACATGTACTAGAGCGTG
TATCGTTTTTTGGGATTAATAATCATATAATCGTTACATGAGCCTTGATACTTTGCTAGA
AATAAGCTCCTCTAAACAAAACCTTCTTTTTAAAAAACACACTTATGTTTTACTTAGTT
TGTGTTGTATCCGAAGTTGATCAACGTTGTAATTTCCACAATAAATCATGACATTTTA
TATGCTCT

>G2519 Amino Acid Sequence (domain in AA coordinates:1-65)
MSHIAVERNRRRQMNEHLKSLRSLTPCFYIKRGDQASIIGGVIEFIKELQQLVQVLESKK
RRKTLNRPSFPYDHTIEPSSLGAATTRVPFSRIENVMTTSTFKEVGACCNSPHANVEAK
ISGSNVVLRVVSRRIVGQVLKIIISVLEKLSFQVLHLNISSMEETVLYFFVVKIGLECHLS
LEELTLEVQKSFVSDEVIVSTN*

>G2520 (133..1197)

AAGGAGTTTTGCATACTCACCAGCCACAATCATTTCTCTCTTCTCTATCTCTCTGGTTT
TGAATCGGCGACGACTGAGTCAACTCGGTGTTGTTACTGGTTTCGTCGTATGTGTTGTAA
CTGATTAAGTTGATGGATCCGAGTGGGATGATGAACGAAGAGGACCGTTAATCTAGCG
GAGATCTGGCAGTTTCCGTTGAACGGAGTTTCAACCGCCGAGATTCTTCTAGAAGAAGC
TTCGTTGGACCGAATCAGTTCGGTGATGCTGATCTAACCACAGCTGCTAACGGTGATCCA
GCGCGTATGAGTCACGCGTTGTCTCAGGCGGTTATTGAAGGTATCTCCGGCGCTTGAAAA
CGGAGGGAAGATGAGTCTAAGTCGGCGAAGATCGTCTCCACCATTGGCGCTAGTGAAGGT
GAGAACAAGACAGAGAAGATAGATGAAGTGTGTGATGGGAAAGCAGAAGCAGAATCGCTA
GGAACAGAGACGGAACAAAGAAGCAACAGATGGAACCAACGAAAGATTATATTCATGTT
CGAGCTAGAAGAGGTCAAGCTACTGATAGTCACAGTTTAGCTGAAAGAGCGAGAAGAGAG
AAAATAAGTGAGCGGATGAAAATCTTGCAAGATCTTGTTCCGGGATGTAACAAGGTTATT
GGAAAAGCACTTGTCTAGATGAGATAATTAATACTATATACAATCATTGCAACGTCAAGTT
GAGTTCTTATCGATGAAGCTTGAAGCAGTCAACTCAAGAATGAACCTGGTATCGAGGTT
TTCCACCCAAAGAGGTGATGATTCTCATGATCATCAACTCAATCTTCTCCATTTTTTTC
ACAAAACAATACATGTTTCTATCGAGGTATTCTCGGGGTAGGAGTCTCGATGTTTATGCG
GTTCCGGTCATTTAAGCATTGCAATAAACGGAGTGACCTCTGTTTTTGCTCCTGCTCCCCA
AAAACAGAACTTAAGACAACTATATTTTCAAAAACATGACATGTTTTCTGTCGATATTCT
CGAGTAGGAGTCCGCTATTAGTTTCATCTAAGCATTGCAATGAACCGTTTGGTCAGCAAGCG
TTTGAGAATCCGGAGATACAGTTCGGGTCGCAGTCTACGAGGGAATACAGTAGAGGAGCA
TCACCAGAGTGGTGCACATGCAGATAGGATCAGGTGGTTTCGAAAGAACGTCTTGA

>G2520 Amino Acid Sequence (domain in AA coordinates: 135-206)

MDPSGMMNEGGPFNLAEIWQFPLNGVSTAGDSSRRSFVGPNGFGDADLTAAANGDPARMS
HALSQAVIEGIGSAWKREDESKSAKIVSTIGASEGENKRQKIDEVCDGKAEAESLGTET
EQKKQMEPTKDYIHRVRRRGQATDSSHSLAERARREKISERMKILQDLVPGCNKVIGKAL
VLDEIINYIQLRQVFEFLSMKLEAVNSRMNPGIEVFPPEVMILMIINSIFSIFFTKQY
MFLSRYSRGRSLDVYAVRSPFKHCNKRSDLCFCSCSPKTELKTTIFSQNMTCFCRYSRVGV
AISSSKHCNEPFGQAFENPEIQFGSQSTREYSRGASPEWLHMQIGSGGFERTS*

>G2533 (1..1080)

ATGATAAGCAAGGATCCAATATCGAGTTTACCTCCAGGGTTTCGATTTTCATCCAACAGAT
GAAGAACTCATTCTCCATTACCTAAGGAAGAAAGTTTCTCTTCCCCAGTCCCGCTTTTCG
ATTATCGCCGATGTCGATATCTACAAATCCGATCCATGGGATTTACCAGCTAAGGCTCCA
TTTGGGGAGAAAGAGTGGTATTTTTTTCAGTCCGAGGGATAGGAAATATCCAACCGAGCA
AGACCAAACAGAGCAGCTGCGTCTGGATATTGGAAAGCAACCGGAACAGATAAATTGATT
GCGGTACCAAATGGTGAAGGGTTTCATGAAAACATTGGTATAAAAAAGCTCTTGTTGTTT
TATAGAGGAAAGCCTCCAAAAGGTGTTAAAACCAATTGGATCATGCATGAATATCGTCTT
GCCGATTCTATCTCCCAAGAATTAACCTCTTCTAGGAGCGGTGGTAGCGAAGTTAAT
AATAATTTTGAGATAGGAATTCTAAGAATATTGATGAGACTGGATGATTGGGTTCTT
TGCCGGATTTCACAGAAATCACACGCTTCATTGTCTATCACCTGATGTTGCTTTGGTCACA
AGCAATCAAGAGCATGAGGAAAATGACAACGAACCAATTCGTAGACCGCGGAACCTTTTG
CCAAATTTGCAAAATGATCAACCCCTTAAACGCCAGAAGTCTTCTTGTTCTGTTCTCAAAC
TACTAGACGCTACAGATTGACGTTTCTCGCAAATTTCTAAGCAAAACCCCGGAAAAT
CGTTCTGAATCAGATTTTTCTTTTCATGATGGCAATTTCTCTAATCTGACATTTACGGA
AACCATTACTTGGATCAGAAGTTACCGCAGTTGAGCTCTCCCACTTCAGAGACAAGCGGC
ATCGGAAGCAAAAGAGAGAGAGTGGATTTTGCGGAAGAAACGATAAACCGCTTCGAAGAAG

ATGATGAACACATATAGTTACAATAATAGTATAGATCAAATGGATCATAGTATGATGCAA
CAACCTAGTTTCTGTAACAGGAACCTCATGATGAGTTCTCACCTTCAATATCAAGGCTAG
>G2533 Amino Acid Sequence (conserved domain in AA coordinates:11-186)
MISKDPISSLPPGFRFHPTEELILHYLRKKVSSSPVPLSIIADVDIYKSDPWDLPAKAP
FGEKEWYFFSPRDRKYPNGARPNRAAASGYWKATGTDKLIAPVNGEGFHENIGIKKALVF
YRGKPPKGVKTNWIMHEYRLADSLSPKRINSSRSGGSEVNNNFGRNSKEYSMRLDDDWVL
CRIYKKSHASLSSPDVALVTSNQHEENDNEPFDVDRGTFLPNLQNDQPLKRQKSSCSFSN
LLDATDLTFLANFLNETPENRSEDFSMIGNFSNPDIYGNHYLDQKLPQLSSPTSETSG
IGSKRERVDFAEETINASKMMNTYSYNNSIDQMDHSMMQQPSFLNQELMMSSHLQYQG*
>G2534 (1..975)
ATGGATAATATAATGCAATCGTCAATGCCACCGGGATTCCGATTTCATCCGACAGAGGAA
GAGCTTGTGGGTATTACCTAGATAGGAAGATCAATTCAATGAAGAGTGCTTTAGATGTC
ATTGTAGAGATTGATCTCTACAAAATGGAGCCATGGGATATACAAGCGAGGTGTAACCTA
GGGTATGAAGAGCAAAAACGAGTGCTACTTCTTTAGTCATAAGGACAGGAAGTACCCTACC
GGGACTAGGACCAACCGAGCCACTGCGGCTGGGTCTTGAAAGCCACGGGTAGAGACAAG
GCGGTACTATCAAAAAACAGTGTCTCGGAATGCGGAAGACACTTGTCTACTACAAGGT
CGAGTCTTAATGAAGAAAGTCCGATTGGATCATGCACGAATACCGTCTCCAAAACCTCC
GAGCTTGCCCCGGTTTCAGGAGGAAGGCTGGGTGGTGTGTCGAGCATTTAGGAAGCCAATT
CCAAACCAGAGGCCATTAGGGTACGAGCCATGGCAGAACCAGCTCTACCACGTGAAAGT
AGTAACAACCTACTCATCTTCAGTGACAATGAACACGAGTCATCATATCGGTGCATCTTCA
TCAAGTCATAACCTTAATCAAATGCTCATGAGCAATAACCACTACAATCCTAATAATACA
TCCTCATCGATGCATCAATATGGCAACATTGAGCTCCCGCAGTTGGACAGCCCGAGCTTG
TCGCCTAGTTTGGGACGAATAAAGATCAGAACGAGAGTTTCGAGCAAGAAGAAGAGAAG
AGCTTTAAGTGTGTGGATTGGAGAACACTAGATACCTTGCTTGAGACACAAGTCATACAT
CCGCATAACCTTAATATTCTTATGTTGAAACGCGAGTCGTATAATCCGGCGCCAAGCTTC
CCTTCCATGCATCAAAGCTATAATGAGGTCGAAGCTAATATTCATCATTCTCTGGATGC
TTCCCTGACTCGTAA
>G2534 Amino Acid Sequence (conserved domain in AA coordinates:10-157)
MDNIMQSSMPPGFRFHPTEELVGYLDRKINSMKSALDVIVEIDLYKMEPWDIQRCKL
GYEEQNEWYFFSHKDRKYPTGTRTNRATAAGFWKATGRDKAVLSKNSVIGMRKTLVYK
RAPNGRKS DWIMHEYRLQNSLAPVQEEGWVVCRAFRKPIPNQRPLGYEPWONQLYHVES
SNYSSSVTMTSHHIGASSSSHNLNQMLMSNNHYNPNNTSSSMHQYGNIELPQLDSPSL
SPSLGTNKDQNESFEQEEKSFNCVDWRTLDTLLETQVIHPNPNILMFETQSYNPAPSF
PSMHQSYNEVEANIHHSLGCFPDS*
>G2573 (34..957)
CCAGATTTAATTTGAGACTCTCAAAGAAACACCATGGAAGAAGAGCAACCTCCGGCCAAG
AAACGAAACATGGGGAGATCTAGAAAAGGTTGCATGAAAGGTAAAGCGGTCAGAGAAC
GCCACGTGTACTTTCCGTGGAGTTAGGCAACGGACTTGGGGTAAATGGGTGGCTGAGATC
CGTGAGCCTAACCGTGGGACTCGTCTCTGGCTCGGCACGTTTAAATACCTCGGTGAGGCC
GCCATGGCTTACGATGAAGCCGCTAAGAAACTCTATGGACACGAGGCTAAACTCAACTTG
GTGCACCCACAACAACAACAAGTAGTAGTGAACAGAACTTGTCTTTTTCTGGCCAC
GGGTGCGGTCTTTGGGCTTATAATAAGAAGCTCGATATGGTTCATGGGTGGACCTTGGT
CTCGGCCAGGCAAGTTGTTACGAGGTTCTTGCTCAGAGAGATCGAGTTTCTACAAGAA
GATGATGATCATAGTCATAATCGATGTTCTGCTTCAAGTGGTTCGAATCTTTGTTGGTTA
TTACCTAAACAAAGTGATTACAAAGATCAAGAGACCGTTAATGCTACGACTAGTTATGGC
GGTGAAGGCGGTGGTGGCTCTACGTTAACGTTTTCGACCAATTGAAACCAAGAATTTG
ATGAGTCAGAATTATGGATTATACAATGGAGCTTGGTCTAGGTTCTTTGTTGGGCAAGAA
AAGAAGACGGAACATGACGTGTCATCGTCTGTTGGATCGTTCGACAACAAGGAGAGTATG
TTGGTTCTAGTTTCGGCGGAGAGAGGATGCATAGGCCGGAGTTGGAAGAGCGAACAGGA
TATTTGGAATGGATGATCTTTTGGAGATTGATGATTAGGTTTGTGATTGGCAAAAAT
GGAGATTTCAAGAATTGGTGTGTGAAGAGTTTCAACATCCATGGAATTGGTTCTGAGAG
TTTTTATTTATTACTATTATTTATCATACATATTTCTTATATTGACTTAGG
>G2573 Amino Acid Sequence (domain in AA coordinates: TBD)
MEEQPPAKKRNRMRGRKGMKGKGGPENATCTFRGVRQRTWGKQWAEIREPNRGRTRLWL
GTFNTSVEAAMAYDEAAKLYGHEAKLNLVHPQQQQQVVNRNLSFSGHSGSGSWAYNKKL
DMVHGLDLGLGQASCSRGSCSERSSFLQEDDDHSHNRCSSSSGSNLCWLLPKQSDSQDQE
TVNATTSYGGEGGGSTLTFSTNLKPKNLMSQNYGLYNGAWSRFLVGQEKTEHDVSSSC

GSSDNKESMLVPSCGGERMHRPELEERTGYLEMDDLLEIDDLGLLIGKNGDFKNWCCEEF
QHPWNWF*

>G2589 (23..1354)

AAAGAAAAGAAAAATAAGATAATGAGGACGAAGACTAAGTTAGTACTCATACCTGATAG
ACACTTTCGGAGAGCCACATTCAGGAAGAGGAATGCAGGGATAAGGAAGAACTCCACGA
GCTGACAACTCTCTGTGACATCAAAGCATGTGCGGTAATCTACAGTCCGTTTCGAGAATCC
AACGGTGTGGCCGTCAACCGAAGGTGTCAAGAGGTGATTCGGAGTTCATGGAGAAGCC
GGCGACAGAACGGTCCAAGACGATGATGAGTCATGAGACTTCTTTCGCGGACCAAATCAC
CAAAGAACAAAACAACTAGAGAGTCTACGTCGTGAAAACCGAGAACTCAGCTTAAGCA
TTTTATGTTTGTGATTGCGTTGGAGGCAAGATGAGTGAGCAACAGTATGGTGCAAGGGACCT
TCAAGATTAAAGTCTTTTACTGATCAATATCTTAATCAGCTTAATGCCAGGAAGAAGTT
CCTTACAGAAATATGGTGAGTCTTCTTCTTCTGTTCTCTCTGTTTGTATGTTGCGGGTGC
CAATCCTCCTGTTGTTGCAGATCAAGCTGCGGTAAGTGTCTCTCTTGTGTTGCTGTTGC
GGGTGCCAATCTTCTGTTGTTGCTGATCAAGCTGCGGTAAGTGTCTCTCTGTTTGC
TGTTGCGGGTGCCAATCTTCTGTTGTTGTCAGATCAAGCTGCGGTTAATGTTCTCTACTGG
ATTTTCATAACATGAATGTGAACAGAAATCAGTATGAGCCGGTTCAGCCCTATGTCCCTAC
TGGTTTTAGTGATCATATTCATATCAGAATATGAAGTCAATCAAAACCAACAAGAGCC
GGTTCATTACCAAGCTCTTGTGTTGCGGGTCCGGTCTTCTATGACTCAGAATCAGTA
TGAGCCCGTTCACTACCAGAGTCTTGTGTCGCGGGTGGCGGTCTTCTATGAGTCAGTT
GCAGTATGAGCCGGTTCAGCCTTATATCCCTACTGTTTTAGTGATAATGTTCAATATCA
GCATATGAATTTGTATCAAAATCAACAAGAGCCGGTTCCTACCAAGCTCTTGGTGTGTC
AGGTGCCGGTCTTCTCTATGAATCAGAATCAGTATGAGCCGGTTCAGCCCTATGTCCCTAC
TGGTTTTAGTGATCATTTTCAGTTTGAGAATATGAATTTGAATCAAAATCAACAGGAGCC
GGTTCAATACCAAGCTCCTGTTGATTTTAATCATCAGATTCAACAAGGAACTATGATAT
GAATTTGAACAGAAATAGATTTGGATCCAAATCAGTATCCGTTTCAAAATGATCCATT
CATGAATATGTTGACAGAAATATCCTTATGAATAAGCGGGTATGTTGAGAGCATGCAC

>G2589 Amino Acid Sequence (domain in AA coordinates: TBD)

MRTKTKLVLPDRHFRRAFRKRNAGIRKKLHELTTLCDIKACAVIYSPFENPTVWPSTE
GVQEVISEFMEKPAERSKTMMSHETFLRDQITKEQNKLES LRRENRETQLKHFMFDCVG
GKMSEQQYGARDLQDLSTFDQYLNQLNARKKFLTEYGESSSSVPLFDVAGANPPVAD
QAAVTVPPLFAVAGANLPVADQAAVTVPPLFAVAGANLPVADQAAVNVPFGFHNMNVN
QNQYEPVQPYVPTGFSDFHIQYQNMNFQNPQEPVHYQALAVAGAGLPMTQNQYEPVHYQS
LAVAGGGLPMSQLQYEPVQPYIPTVFSQNVQYQHMLNLYQNQEPVHYQALGVAGAGLPMN
QNQYEPVQPYVPTGFSDFHQQFENMNLNQNPQEPVQYQAPVDFNHQIQQGNYDMNLNQNM
LDPNQYPFQNDPFMNLTEYPYE*

>G2687 (45..1139)

CTCTGTCTCTCGTATCTTTCTACTACTCTGTTTCTTGAATTCTAATGAACAACATCGACG
ACGCAAGACGGAGACTTCAGTGTCTTCAGGTTCAAGCGACTCTTCTTGCCTCTCAAGA
AACGCTAGAGACTTGATGACGAACCAAGCAACCGCCCTAGTGGTTTCTGCTTTCACCAAGA
CGGTTGTGGCTTCTGGCAATGTCAAGTACAAAGGAGTCGTTTCAGCAACAGAACGGTCATT
GGGGTGCCAGATTTACGCAGACCACAAAAGGATTTGGCTTGGAACCTTCAAATCCGCTG
ATGAAGCCGCCACGGCTTACGATAGTGCATCTATCAAACCTCCGAAGCTTTGACGCTAACT
CGCACCGGAACCTTCCCTTGGTCTACATCACTCTCAACGAACAGACTTTCAAATTTGCT
ACACAACAGAGACTGTGTTGAACATGATCAGAGACGGTTCGTACCAACACAAATTCAGAG
ATTTTCTCAGAATCAGATCTCAGATTGTTGCGAGTATCAACATCGGGGGACCAAAACAAG
CCCGAGGAGAAGTGAATCAAGAATCAGACAAGTGTTTTCTTGACACAGCTTTTTCAGA
AGGAATTGACACCGAGCGATGTAGGGAACTAAATAGGCTTGTGATACCTAAAAAGTATG
CAGTGAAGTATATGCTTTTCATAAGCGCTGATCAAAGCGAGAAAGAAGAGGGTGAAATAG
TAGGATCTGTGGAAGATGTGGAGGTTGTGTTTTACGACAGAGCAATGAGACAATGGAAGT
TTAGGTATTGTTACTGGAAGAGTAGCCAGAGCTTTGTCTTACCAGAGGATGGAATAGTT
TCGTGAAGGAGAAGAATCTCAAGGAGAAGGATGTTATTGCCTTCTACACTTGCAGATGTC
CGAACAATGTGAAGACATTAGAAGGTCAAAGAAAGAACTTCTTGATGATCGATGTTTCATT
GCTTTTCAGACAACGGTTCCGTTGAGTCTGAGGAAGTAAGTATGACGGTTCATGACAGTT
CAGTGCAAGTAAAGAAAACAGAAAACCTTGGTTAGCTCCATGTTAGAAAGATAAAGAAACCA
AATCAGAGGAGAACAAAGGAGGGTTTATGCTGTTTGGTGTAAGGATCGAATGTCCTTAGG
GAATTTTCTTTAAAGTTTCTTACTTCAACTAGAACTTGTTTTACTTGTAACCT

>G2687 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNIDDAKTETSVSSGSSDSFLPLKKRMRLDDEPENALVSSSPKTVVASGNVYKGVVQ
QQNGHWGAQIYADHKRIWLGTFKSADEAATAYDSASIKLRSFDANSHRNFPWSTITLNEP
DFQNCYTTETVLNMIRDGSYQHKFRDFLRIRSQIVASINIGGPKQARGEVNQESDKCFSC
TQLFQKELTPSDVGKLNRLVIPKYAVKYMPIFISADQSEKEEGEIVGSVEDVEVVYFDRA
MRQWKFRYCYWKSSQS FVFTRGWNSFVKEKNLKEKDVI AFYTCDPNNVKTLEGQRKNFL
MIDVHCFSDNGSVVAEEVSMTVHDSSVQVKKTENLVSSMLEDKETKSEENKGGFMLFGVR
IECP*

>G27 (83..622)

CAAAATACCAAAAAACAAACATTTTTTTTAACTCTCCCAACATTTTTTTCTCTTTCTCT
CGTTACATTAAATTATCTTTAGATGCAAGACTCTTCTCTCACGAATCGCAACGTAACTT
CCGGTCACCGGTGCGCGAGAAAACCGGAAAGAGTTCTAAGACTAAAAATGAGCAAAAAGG
TGTTTCTAAACAACCAATTTTCGTGGGGTCAGAATGAGACAATGGGGAAAAATGGGTGTC
TGAAATTAGAGAACCAAGAAAGAAATCAAGAATATGGCTCGGTACTTTCTCTACGCCGGA
GATGGCGGCGCGTGCACACGACGTGGCGGCTTTAGCCATCAAAGGTGGCTCTGCCACCT
TAATTTCCCGGAGCTAGCTTACCATTGCGGAGACCGGTAGCGCGGACCCCTAAAGACAT
TCAAGAAGCCGCGCGCAGCAGCTGCGGTTGACTGGAAAGCACCGGAGTCTCCGTCTAG
CACCGTGACGTCTCTCCAGTCCGCGACGACGCTTTCTCCGATCTTCTGATCTTTTGCT
TGACGTGAATGATCACAACAAAACGATGGATTCTGGGACTCGTTCCGTACGAAGATCC
TTTCTTCTTGAAAAATTACTAGAAGGCAAATCTTGCCGGCGAACGGATTTTCCGGTGGT
TTCCCGGTAAATAAGAAGACGATGTCGTTTGTACCTTTTTTGTCTACGATGGGAAATTT
CTTTTTTTTTTACGTGTGAGTAAAGTTTCCGAATGTGTGATGTGTAAGTAAGTACAGGT
TATTTAATTTCTTTTTTTTTGTACAAATACGTACGTATTACCAAAAAGTTTTCATTTATT
GTGCTTTTATCTTCCAAATTCATTAAAAA

>G27 Amino Acid Sequence (domain in AA coordinates: 37-104)

MQDSSSHESQRNLRSPVEKTKGSSKTKNEQKGVSKQPNFRGVRMRQWGKVVSEIREPRK
KSRIWLGTFSTPEMAARAHDVAALAIKGGSAHLNFPDELAYHLPRPASADPKDIEAAAAA
AAVDWKAPESPSTVSSPVADDAFSDLPDLLLLDVNDHNKNDGFWDSPFYEDPFFLENY*
>G2720 (1..894)

ATGGAAGCGAAGAAGGAAGAGATAAAGAAAGGTCCATGGAAAGCCGAAGAAGACGAAGTA
CTCATCAACCATGTCAAGAGATACGGTCCTCGTGATTGGAGCTCCATTTCGATCCAAAGGT
CTTCTTCAACGCACCGGCAAATCCTGTCTCTCGTTGGGTCAATAAACTCCGTCCCAAT
CTCAAAAATGGATGCAAGTTCTCGCTGACGAAGAGAGGACTGTGATTGAGTTACAATCT
GAGTTTGGTAACAAATGGGCGAGAATCGCTACGTATCTACCGGGAAGAAGTGAACGAT
GTGAAGAATTTCTGGAGTAGCAGACAAAAGAGACTCGCTAGGATTCTTCATAACTCCTCT
GATGCATCGAGTTTCGAGTTTCAATCCCAATCTTCTTCTCTCATCGACTCAAGGGCAA
AACGTCAAACCAATCCGTCAATCCTCTCAGGGTTTTGGTTTGGTTGAGGAAGAGGTTACA
GTTTCTTCTTCTCATGTTCCAGATGGTTTCTTATTCTATCTGATCAAGTTGGTGATGAAGTC
TTGAGGTTGCCGATTTGGGTGTTAAGTTAGAGCATCAGCCTTTTCGCTTTTGGCACTGAT
CTTGTCCTAGCAGAGTACTCTGACTCACAGAATGATGCAAATCAGCAAGCAATCAGCCCT
TTCTCTCAGAAAGCAGAGAGCTTTTGGCTAGACTTGACGACCTTTTACTATGATATA
CTTGGACCACTGATTCTTCTGAGCCATTGTTGCTCTCCCTCAGCCGTCTTTCGAGCCT
TCGCCTGTGCCGAGAAGATGCAGACATGTTTCAAAGGATGAAGAAGCTGATGTTTTCTTA
GACGATTTCCAGCTGACATGTTTGATCAGGTTGATCCAATCCCAAGTCCTTAG

>G2720 Amino Acid Sequence (domain in AA coordinates: 10-114)

MEAKKEEIKKGPWKAEEDEVLINHVKRYGPRDWSSIRSKGLLQRTGKSCRLRWVNLKRPN
LKNGCKFSADEERTVIELQSEFGNKWARIATYLPGRTDNDVKNFWSSRQKRLARILHNSS
DASSSFFNPKSSSSHRLKGKVKPIRQSSQGFGLVEEVTVSSSCSQMVPYSSDQVQDEV
LRLPDLGVKLEHQPFAGTDLVLAEYSDSQNDANQQAISPFSPESRELLARLDDPFYDI
LGPADSSEPLFALPQPFEPSPVPRRHRVSKDEEADVFLDDFPADMFDQVDPISP*

>G2787 (142..1584)

TCTCAGAGCAAAAAACAAAAAAGAAAAAACCCTAAATCTAAATCTCACCTTCCA
CCTCTGTCTTTTTTTTTTTGTTCTTTTTTTTTTTTACTGTATCTTCTCTCTTTG
CTCTGCAAAAATCTCACATCCATGGATCCATCTCTTGGTGATCCTCATCATCCTCCTCAG
TTCACCCCTTTTCTCATTTTCCCACCTCAAATCATCATCTTTAGGACCAAATCCGTAC
AATAACCATGTCGCTTCCAACCGCAGCCGCAACGCAACGCAAAATCCCGCAACCGCAG
ATGTTTCAGTTATCTTCCACATGTTCAATGCCCAACCTCCTTACTCCGAAATGATTTGC
GCTGCGATTGCGGCGTTAAACGAACCGGATGGTTCGAGCAAGATGGCAATTTTCGAGATAC

ATCGAGAGATGTTACACCGGTTTAACTTCTGCTCATGCTGCTTTGTTGACTCACCATCTC
 AAGACTTTGAAGACCAGTGGTGTCTTTCTATGGTTAAGAAATCTTACAAAATTGCTGGT
 TCTTCTACTCCTCCTGCTAGTGTAGCTGTGCTGCTGCTGCCGCCGCTCAAGGTCTCGAT
 GTTCCCAGATCTGAGATTCTCCATTCAAGTAACAACGATCCCATGGCTTCTGGCTCTGCT
 TCTCAGCCTCTGAAACGAGGTCGTGGTCGTCCTCTAAGCCTAAACCTGAATCTCAACCA
 CAACCACTACAGCAACTTCCACCGACCAATCAAGTCCAGGCTAACGGACAGCCAATCTGG
 GAACAGCAGCAAGTTCAATCACCTGTTCCGGTTCGACTCCGGTTACAGAGTCGGCGAAG
 AGAGGACCTGGTCGTCCAAGGAAGAACGGTCTGCTGCTCCTGCTACTGCACCAATCGTT
 CAAGCTTCGGTTATGGCTGGAATTATGAAACGTAGAGGTAGACCACCGGGTCGTGAGCT
 GCTGGGAGACAGAGGAAGCCCAAATCCGTTTCTTCTACTGCCTCTGTGTATCCTTATGTT
 GCTAATGGTGTAGACGCAGAGGAAGGCCTAGGAGAGTTGTTGACCCTAGCAGTATTGTT
 AGTGTGCTCCAGTAGGTGGTGAAATGTGGCAGCGGTTGCGCCAGGGATGAAGCGTGGA
 CGTGACGACCACTAAGATTGGTGGTGTATCAGTAGGCTTATTATGAAGCCTAAGAGA
 GGACGAGGACGTCCTGTAGGTAGACCCAGAAAGATTGGAACATCAGTCACGACTGGGACA
 CAAGATTCTGGAGAACTCAAGAAGAAGTTTGATATTTTCAAGAGAAAGTGAAAGAAAT
 GTGAAGGTGTTGAAGGATGGAGTTACAAGTGAGAATCAAGCAGTGGTGCAAGCCATAAAA
 GATCTGGAAGCACTAACAGTACGGAGACCGTTGAGCCACAAGTTATGGAAGAAGTGCAG
 CCAGAGGAGACTGCAGCACCACAGACTGAAGCTCAACAACTGAAGCTGCTGAGACACAA
 GGAGGACAAGAAGAAGGACAAGAAAGAGAAGGAGAAACACAGACCCAGACAGAAGCAGAG
 GCAATGCAAGAAGCTCTGTTCTGAAGAATAATAATGATCTAGAAAACAACCTAGACATAA
 TAGCCTTGTTGTTGGCGTTAGGAGTGTGTTTTTTTAGTTGTTTTAGGTGTTGGAATCGC
 ATCTTAAATTATATAAAAATCTATAAGGAATTTAATTTTTCTAGGTTTGTGCTGCA
 GAAGAAGAAATAGTAGACTCGTTAATGGTGTGTTGTCGGTGTGCTTTAACCAAACCAT
 AAGACGTGGCTGTAAATTAGCGATGTTTCTAGTCTTCCATCTTTAATAATCTCTTATTGC
 GTCTGTGCCTTTGTTTTT

>G2787 Amino Acid Sequence (domain in AA coordinates: 172-192, 226-247, 256-276, 290-311, 245-366)

MDPSLGDPHHPQFTFPFHPFPTSNHHPLGPNPYNHVVFPQPPQTQTQIPQPMFQLSPH
 VSMPPHPYSEMICAIAALNEPDGSSKMAISRYIERCYTGLTSAHAALLTHHLKTLKTSG
 VLSMVKKSYKIASSTPPASVAVAAAAAQGLDVPREILHSSNNDPMASGSASQPLKRG
 RGRPPKPKPESQPQPLQQLPPTNQVQANGQPIWEQQVQSPVPVPTPTVESAKRGPGRPR
 KNGSAAPATAPIVQASVMAGIMKRRRPPGRRRAAGRQRKPKSVSTASVYPYVANGARRR
 GRPRRVDPSSIVSVAPVGGENVAAPVPMKRRGRPPKIGGVISRLIMKPKRGRGRPVG
 RPRKIGTSVTTGTQDSGELKKKFDIFQEKVKEIVKVLKDVTSSENQAVVQAIKDLEALTV
 TETVEPQVMEEVQPEETAAPQTEAQQTAAETQGGQEBEQEREGETQTQTEAEMQEALEP
 *

>G2789 (82..879)

CTTTAGGGACACCAATCTATTCAACCTAAAAGCCTTCTTTTCCCTATATTGACCAACT
 TTTTAGCGAATCAGAAGAGGAATGGATGAGGTATCTCGTTCTCATACACCGCAATTCTA
 TCAAGTGATCATCAGCACTATCACCATCAAAACGCTGGACGACAAAAACGGGCAGAGAA
 GAAGAAGGAGTTGAACCAACAATATAGGGGAAGACCTAGCCACCTTTCTTCCGGAGAA
 GAGAATATCAAGAAGAGAAGGCCACGTGGCAGACCTGCTGTTCCAAGAACAACCCAAA
 GCACCAATCATAGTCACTCGCGACTCCCGGAACGCTTCAGATGTCACGTATGGAGATA
 ACCAACGCTGCGATGTAATGGAAGCCTAGCCGTCTTCGCTAGACGCCGTGAGCGTGGC
 GTTTGCGTCTTGACCGGAACGGGGCGTTACAAACGTCACCGTTAGACAACCTGGCGGA
 GCGTCGTCAGTTTACACGGACGGTTTGAGATTCTTCTCTCTCGGGTTCGTTTCTTCTCT
 CCACCGGCACCAACAGCTGCGTCTGGTTTAAAGGTTTACTTAGCCGGTGGTCAAGGTCAA
 GTGATCGGAGGCAGTGTGGTGGGACCGCTTACGGCATCAAGTCCGGTGGTCTGTTATGGCA
 GCTTCATTTGGAACGCATCTTACGAGAGGCTGCCACTAGAGGAGGAGGAGAACTGAA
 AGAGAAATAGATGGAACCGCGGCTAGGGCGATTGGAACGCAACGCAGAAACAGTTAATG
 CAAGATGCGACATCGTTTATTGGGTCGCCGTCAATTTAATTAACCTCTGTTTCGTTGCCA
 GGTGAAGCTTATTGGGGAAACGACACCGCTCTTCTAAGATAATATCATTGATAATATA
 AGTTTCGTCTTCTTATCTTTTTTCACTTTTTTCACTTTCTTAGGTTTGTGTTTT
 AACGTTTGATTAATACCTGAAGGTTTTTGGAAAATTTTCGATCGGATAAAGGATTATG
 TTGCGAGCCGAAACGCGGCC

>G2789 Amino Acid Sequence (domain in AA coordinates: 53-73, 121-165)
 MDEVSRSHTPQFLSSDHQHYHHQNAQRQKRGREEGVENPNNIGEDLATFPSGEBENIKRR

PRGRPAGSKNPKAPIIVTRDSANAFRCHVMEITNACDVMESLAVFARRRQRGVCLTGN
GAVTNVTVRQPGGGVSLHGRFEILSLSGSFLPPPAPPAASGLKVYLAGGQGVIGGSV
GPLTASSPVVMAASFGNASYERLPLEEEEEETEREIDGNAARAIGTQTQKQLMQDATSFI
GSPSNLINSVSLPGEAYWGTQRPSF*

>G31 (13..615)

CTTTTATAAGCAATGGCTCCAAGACAGGCGAACGGTAGAAGCATTGCCGTGAGTGAAGGC
GGCGGAGGGAAGACGATGACGATGACGACGATGCGGAAGGAAGTGCACCTTTAGAGGTGTG
AGGAAGCGTCCATGGGGTAGATACGCGGCGGAGATCCGTGACCCGGGAAAGAAAACCCGG
GTTTGGCTCGGGACATTTCGACACGGCGGAGGAAGCTGCAAGAGCTTACGACACCGCCGCT
AGAGAGTTTCGTGGCTCCAAAGCAAAGACTAATTTCCCTCTTCCCGGAGAGTCTACTACG
GTTAACGACGGTGGCGAGAACGATTCTTACGTCAACCGTACGACGGTGACGACGGCGCGT
GAGATGACGCGTCAGAGATTTCCGTTTGCATGTCAACCGGAGCGTAAAGTCGTGGTGGT
TATGCTTCTGCTGGTTTTTTCTTCGATCCGTCAGAGCTGCTTCGTTAAGAGCAGAGCTT
TCTCGGGTTTGTCCGGTTCCGGTTTGATCCGGTTAATATCGAGTTGAGTATTGGTATTCTGA
GAAACCGTAAAAGTTGAACCGAGAAGAGAATAAACCCTGGATCTTAACCTAGCTCCACCG
GTGGTGGACGTTTAGATTTTTTCTTCTTTTCATAATTTGTATTTTACATTGCCGGAATA
TAATTAATGTTTTCTTTAG

>G31 Amino Acid Sequence (domain in AA coordinates: TBD)

MAPRQANGRSIAVSEGGGKMTMTTMRKEVHFRGVKRPWGRYAAEIRDPGKKTRVWL
TFDTAEAAARAYDTAAREFRGSKAKTNFPLPGESTTVNDGGENDSYVNRITVTAREMTR
QRFPFACHRRERKVVGGYASAGFFDPSRAASLRAELSRVCPVRFPVNIELSIGIRETVK
VEPRRELNLDLNLAPPVVDV*

>G33 (20..757)

ATTCTCCCCCAACCAAAATATGACCACAGAAAAAGAGAATGTCACTACGGCCGTGGCCGT
GAAAGACGGCGGAGAAAAGAGTAAGGAAGTGAGTGACAAGGGCGTAAAGAAGAGAAAGAA
TGTAACCTAAGGCCCTGGCCGTGAATGACGCGGAGAAAAGAGTAAGGAAGTGCCTTACAG
GGGTGTAAGGAGGAGACCATGGGGGAGATATGCTGCGGAGATCCGTGATCCGGTAAAGAA
AAAACGGGTCTGGCTCGGGTCTTCAACACGGGGGAGGAAGCCGCCAGAGCCTACGACTC
CGCTGCCATAAGGTTTCGAGGATCGAAAGCTACTACTAATCTCCCTCTAATCGGATACTA
TGCGGATTTCTTCGCGCAGCGCCGTGAACAACAACCTTTCCGAGACGGTGAGTGATGGAAA
TGCCAACCTCCCTCTCGTTGGAGACGATGGGAATGCTTTGGCTTCTCCGGTGAACAACAC
CCTTTCCGAAACGGCGCGTGATGGAACACTTCCATCGGATTGTACGACATGTTATCTCC
GGGGGTGGCTGAAGCGGTTGCTGGATTTTTCTTAGATCTGCCTGAAGTTATTGCGTTGAA
AGAGGAGCTTGATCGAGTTTGTCTGACAGTTTGTAGTCCATTGATATGGGGTTGACTAT
TGGTCTCTCAAACCGCCGTGGAGAGCCTGAGACTTCTCCGCCGTGGATTGTAAGCTGCG
AATGGAACCGGATCTTGACCTCAACGCAAGTCCCTAAAGATTGATCTGATGTTGTTGTCC
TTGAATAAGTTTGTATCTTGTGCTCTTCTGATTGTCTGTACTTCTATTGGTTGATTGCG
TGCTTTTGGAGACAAAACAAACATTTTTTTATGTATTAAAAAAGGTAATTGAACTATT
ATCGAAAAA

>G33 Amino Acid Sequence (domain in AA coordinates: 50-117)

MTTEKENVTTAVAVKDGGEKSKEVSDKGVKKRKNVTKALAVNDGGEKSKEVRYRGVRRRP
WGRYAAEIRDVPVKKRVWLGSFNTGEEAARAYDSAAIRFRGSKATTNPLIGYYGISSAT
PVNNNLSETVSDGNANLPLVGDDGNALASPVNNTLSETARDGTLPSDCHDMLSPGVAEAV
AGFFLDLPEVIALKEELDRVCPDQFESIDMGLTIGPQTAVEEPETSSAVDCKLRMEPDLD
LNASP*

>G342 (1..723)

ATGGACGTCTACGGCATGTCTTCACCGGACTTGCTTCGTATCGACGACCTTCTCGATTTC
TCCAACGACGAAATCTCTCTCTCTCCACCGTCACTTCTCCGCCGCTTCTCTCCGCC
GCTTCTTCCGAAAACCTTTCAGCTTTCTCTCTTCCACCTACACTTCTCTACTCTCTC
ACCGACTTCACTCACGATCTCTGCGTTCCAGTGACGACGAGCTCATCTCGAATGGTTA
TCGCGATTGCTTGACGATTCACTCTCCGATTTCAGCAAAATCTTTAACCATGACCGTT
AGACCCGAGATTTCACTACCGGAAAACCTAGAAGTCGCCGATCAAGAGCACCAGCACCT
TCCGTAGCTGGAACCTGGGCTCCGATGTCTGAATCAGAGCTTTGTCACTCCGTCGCTAAA
CCTAAACCGAAGAAAGTCTACAACGCTGAATCGGTTACGGCGGATGAGCGGAGCGGTGC
ACGCACTGTGCCTCGGAGAAAACGCCACAGTGGAGAACTGGACCGCTTGGACCTAAAACA
CTTTGTAAACGCTTGTGGAGTTCTGTTACAAATCAGGGAGGCTTGTACCGGAATACAGACCG
GCGTCGAGTCCGACGTTTGTATTGACTCAGCACTCGAACTCTCATCGGAAAGTTATGGAG

CTCCGGCGACAGAAGGAACAACAAGAATCTTGC GTTCGAATCCGCCGTTTCAGCCGCAG
TAA

>G342 Amino Acid Sequence (domain in AA coordinates: 155-190)
MDVYGMSSPDLLRIDDLLDFSNDEIFSSSSTVTSSAASSAASSENPFSPSSSTYTSPTLL
TDFTHDLCPVSDDAAHLEWLSRFVDDSFSDFPANPLTMTVRPEISFTGKPRSRRSRAPAP
SVAGTWAPMSESELCHSVAKPKPKKVYNAESVTADGARRCTHCASEKTPQWRTGPLGPKT
LCNACGVRYKSGRLVPEYRPASSPTFVL TQHSNSHRKVMELRRQKEQQESCVRIPPFQPO
*

>G352 (80..817)
AATACACCACACACTTCACTCTTCTTCATCTTCTTCTTCTTAAATAGCTCGAAATCACA
TCTCACAGAATTAAATCTTATGGCTCTCGAGACTCTCAATTCTCCAACAGCTACCACCAC
CGCTCGGCCTCTTCTCCGGTATCGTGAAGAAATGGAGCCTGAGAATCTCGAGCAATGGGC
TAAAGAAAACGAACAAAACGTCACCGTTTTGATCACGGTCATCAGAATCAAGAAACGAA
CAAGAACCTTCTCTTGAAGAAGAGTATCTCGCTCTTTGTCTCCTCATGCTCGCTCGTGG
CTCCGCGGTACAATCTCTCTCTCTCTCTCTCTACCGTCACGTGCGTCACCGTCCGATCA
CCGAGATTACAAGTGTACGGTCTGTGGGAAGTCCTTTTCGTCATACCAAGCCTTAGGTGG
ACACAAGACGAGTCACCGGAAACCGACGAACACTAGTATCACTTCCGGTAACCAAGAACT
GTCTAATAACAGTCACAGTAACAGCGGTTCCGTTGTTATTAACTTACCGTGAACACTGG
TAACGGTGTAGTCAAAGCGGAAAGATTACACTTGCTCAATCTGTTTCAAGTCGTTTGC
GTCTGGTCAAGCCTTAGGTGGACACAAACGGTGTCACTATGACGGTGGCAACACGGTAA
CGGTAACGGAAGTAGCAGCAACAGCGTAGAACTCGTCTGGTAGTGACGTACGCGATGT
TGATAATGAGAGATGGTCCGAAGAAAGTGGGATCGGTGGCCACCGTGGATTGACCTAA
CTTACCGGCTGATCAAGTCTCAGTGACGACTTCTTAA

>G352 Amino Acid Sequence (domain in AA coordinates: 99-119,166-186)
MALETLSNPTATTTARPLRLRYREEMEPENLEQWAKRKRTKRQRFDHGHQNETNKNLPSE
EEYLALCLLMLARSAVQSPPLPPLPSRASPSDHRDYKCTVCGKSFSSYQALGGHKTSHR
KPTNTSITSGNQBSNNSHNSGVSVINVTNTGNGVSQSGKIHTCSICFKSFASGQALG
GHRCHYDGGNNGNGNGSSNSVELVAGSDVSDVDNERWSEESAIGGHRGFDLNLPAQV
SVTTS*

>G357 (1..615)
ATGCAGAACAAACACAAATGCAAGCTCTGTTCCAAGAGTTTCTGTAATGGCAGAGCACTT
GGTGGTCACATGAAGTCTCACTTGGTCTCATCTCAGTCTTCAGCTCGGAAGAACTAGGT
GACTCGGTCTATTCTTCTTCTTCTTCTTCTTCCGATGGTAAAGCGCTCGCCTACGGGTTA
CGAGAGAACCCGAGGAAGAGTTTCCGGGTCTTAAATCCGGATCCTGAGTCATCCACAATT
TACAACAGTGAGACAGAGACCGAACCTGAATCCGGAGACCCGGTTAAGAAACGGGTGAGA
GGAGATGTTTCAAAGAAGAAGAAGAAGGCAAAGAGTAAGAGAGTGTTTGAGAACTCG
AAGAAGCAAAAGACAAATTCACGAGTCACCAGAACAGCAGAGTCTGTCTCTGATGGTTCT
CCTGAACAAGATTTAGCTATGTGCTTGATGATGCTGTCAAGAGATTCAAGGGAGCTCGAG
ATTAAGTGAAGAAACCGGAGGAAGAGAGGAAAGCCGGAAGAAAGACATTTCCCTGAGCTC
CGTCTGTATGATAGATCTGAATCTTCTCCGCCGCAAGAAGCTGAAGCTGTCACCGTC
GTTTCAGCCATATAA

>G357 Amino Acid Sequence (domain in AA coordinates: 7-29)
MQNKHKCKLCSKSFENGRALGGHMKSHLVSSQSSARKKLGDVSYSSSSSSSDGKALAYGL
RENPRKSFRVFNPDPESSITYNSETETEPESGDPVKRVRGDSVSKKKKKAKSKRVFENS
KKQKTIHESPEPASSVSDGSPEQDLAMCLMMLSRDSRELEIKLKKPEEERKPEKRHFPEL
RRCMIDLNLPPQAEAEAVTVVSAI*

>G358 (1..855)
ATGGGTCAAGATGAGGTTGGGAGTGATCAGACGCAAAATCATAAAGGGAAACGTACGAAG
CGACAAAGATCGTCTTCGACGTTTGTGGTGACGGCGGCGACAACAGTGACTTCAACAAGT
TCATCGGCCGGTGGAAGTGGAGGAGAAAGAGCTGTTTCAGATGAATACAACCTCGGCGGTT
TCGTCTCCGGTGACTACTGATGTGACGCAAGAAGAAGAAGACATGGCGATTGTCTCATC
ATGTTAGCTCGTGGGACAGTTCTTCCATCGCCGGATCTCAAGAACTCGAGAAAAATTCAT
CAGAAGATTTCTGTCGAGAAATCTAGTTTCTATGTGTACGAGTGTAACCGTGTAACCGG
ACGTTTTCTGTCGTTCCAAGCACTTGGTGGACACAGAGCGAGCCACAAGAAGCCGAGGACG
TCGACTGAGGAAAAGACTAGACTACCCCTGACGCAACCCAAGTCTAGTGCATCAGAAGAA
GGGCAAAACAGTCATTTCAAAGTTTCCGGCTCAGCCCTAGCTTTCACAGGCAAGTAACATC
ATCAACAAGGCAACAAAGTACACGAGTGTTCATCTGCGGTTCTGAGTTCACTTCCGGG

CAAGCTCTCGGTGGTCACATGAGGCGGCACAGGACAGCCGTAACCACGATTAGCCCCGTT
GCAGCCACCGCAGAAGTAAGCAGAAACAGTACAGAGGAAGAGATTGAGATCAATATAGGC
CGTTCGATGGAACAGCAGAGGAAATATCTACCGTTGGATCTTAATCTACCAGCACCAGAA
GATGATCTAAGAGAGTCAAAGTTTCAAGGGATAGTATTCTCAGCAACACCAGCGTTAATA
GATTGTCATTACTAG

>G358 Amino Acid Sequence (domain in AA coordinates: 124-135, 188-210)
MGQDEVGSDQTQIIKGRKTRQRSSSTFVVTAAATTVTSTSSAGSGGERAVSDEYNSAV
SSPVTTDCTQEEEDMAICLIMLARGTVLPSPDLKNSRKIHQKISSENSSFVVECKTCNR
TFSSFQALGGHRASHKKPRTSTEEKTRLPLTPKSSASEEQNSHFKVSGSALASQASNI
INKANKVHECSICGSEFTSGQALGGHMRRHRTAVTTISPVAATAEVSERNSTEEIEINIG
RSMEQQRKYLPLDLNLPAPEDDLRESKFQGI VFSATPALIDCHY*

>G360 (1..543)
ATGTGGAACCCCTAACAAATTTGAAGAATTGGAGGATGATGATGAATCTTGGGAAGTCAAA
GCCTTTGAGCAAGACACTAAAGCAACATCTCTGGTACCCTTGGCCCTCAAGATCTTAC
ACTTGCAATTTCTGCCGCCGTGAGTTCGGTCTGCTCAAGCCTTAGGCGGTACATGAAT
GTCCACCGCCGTGACCGCGCCTCATCTAGGGCTCATCAAGGTTCCACCGTTGCCGGCTGCC
GCTAGAAGCGGCCACGGGGGATGTTACTCAATTCTTGTGCTCCGCCGTTGCCCTACAACG
ACACTTATAATACAATCCACGCGAGTAACATTGAAGGTTTGTCCATTCTACCAACTG
CAAAACCCTAGTGGCATTTTTGGTAATTCTGGTGACATGGTGAATCTTTATGTAGAAGTT
CCTCCTCGGCTTATTGAATATTCGACAGGAGATGATGAGAGCATTGGCTCGATGAAAGAA
GCGACAGGAACATCAGTGGATGAGCTTGATCTTGAACCTCGGCTAGGGCACCATCCACCG
TGA

>G360 Amino Acid Sequence (domain in aa coordinates: 42-62)
MWNPNKIEELEDDESWEVKAFQDTKGNISGTTWPPRSYTCNFCRREFRSAQALGGHNM
VHRRDRASSRAHQGSTVAAAARSGHGGMLNNSCAPPLPTTLIIQSTASNIEGLSHFYQL
QNPSGIFGNSGDMVNLYVEVPPRLIEYSTGDDESIGSMKEATGTSVDELDLELRLGHHP
*

>G362 (195..830)
ATAAAAAACCCCTTCATACAATATAAAATTTCTTTAGACATACAATATATTATACTATTAC
AGATGCAATGCATCATTAGTTACAAACTATTAACTAAATATCCCCCGTCTCTCTCTTGC
TATATAAAGAAGATCATTACACATCTCCTTAAGCAAATTAAACCCATCGATAAACACAT
ACGTTACACATATATGTCTATAAATCCGACAATGTCTCGTACTGGCGAAAGTTCTTCAG
GTTTCGTCTCCGACAAGACGATAAAGCTATTTCGGCTTCGAACATCATCAGCGGCAGTCGTA
CGCCGGAAATCACGACGCGGAAAGCGTGAGCTCGTCCACAAACACGACGTCGTTAACAG
TGATGAAAAGACACGAGTGGCAATACTGCGGTAAAGAGTTTGCAAATTCTCAAGCCTTAG
GAGGTACCAAAAACGCTCACAAGAAGGAGAGGTTGAAGAAGAAGAGGCTTCAGCTTCAAG
CTCGCGAGCCAGCATCGGCTATTATCTACCAACCACCAACAACCCATAACGACGTCAT
TTCAGAGACAATACAAAACGCCGTCGTATTGTGCATTCTCTCCATGCACGTGAATAATG
ATCAGATGGGTGTGTACAACGAAGATTGGTCGTGAGGTCGTGCGAGATTAACCTCGGTA
ATAATGACACGTGCCAAGATCTTAATGAACAAAGCGGTGAGATGGGTAAGCTGTACGGTG
TTCGACCGAACATGATTCAGTTCAGATCCAGAGATCTGAGTTCTCGTTCTGATCAGATGAGAA
GTATTAACCTCGCTGGATCTTCATCTAGGTTTTGCGGAGATGCGGCATAACAAATTAAG
AGAGATATATGATTAAGATTATATGTACTATAGTGGCGTATTTCAATTGGGATCATGAAGG
GGAAAAACGAGACATATAGTATTCTTGATGCAATTTGAGTTTTGTAATTTATTAGGTT
TATGTATGTTTTCGAAG

>G362 Amino Acid Sequence (domain in AA coordinates: 62-82)
MSINPTMSRTGESSSGSSDKTIKLFGEFELISGSRTPETITAESVSSSTNTTSLTVMKRH
ECQYCGKEFANSQALGGHQNAHKERLKKRLQLQARRASIGYYLTNHQQPIITTSFQROY
KTPSYCAFSSMHVNDQMGMVYNEDWSSRSQINFNNDTQCQLNEQSGEMGKLYGVRPNM
IQFQRLSSRSQMRINSLLDLHLGFAGDAA*

>G364 (64..516)
AAGCTTGATATCGCCTCTCTCTAATCTCTCTTCTCTCTATCTCTAAGAATATATAAA
GGTATGGACTACCAGCCAAACACATCCCTACGTCTAAGCCTACCAAGTTACAAGAACCAC
CAACTAAACCTAGAAGTTGTTCTCGAGCCTTCTTCATGTCTTCTTCTTCTTCTTCC
ACGAACCTCATCATATGTTTGGAGCAGCCTAGGGTATTCTCATGTAACCTATTGTCAAAGA
AAGTTTTACAGCTCTCAAGCTCTTGGTGGTCATCAAAACGCTCATAAGCTTGAGAGAACC
TTAGCCAAGAAGAGTCGAGAAGCTCTTTAGATCCTCAACACTGTTGATTCTGATCAGCCT

TACCCGTTCTCCGGTCGCTTTGAGCTTTACGGCCGTGGCTACCAAGGATTTCTCGAAAGT
GGCGGCTCGAGGACTTCTCCGCCCGCCGTGTGCCGAGAGTGGTCTTGATCAGGATCAG
GAGAAGAGTCACCTTGACTTATCCTTAAGGCTCTAAAAGAATCTTATATTTTGTAGTCT
ATATATTATCATATCAATTGTTAATCTTAAAATTGATTGTTTACTTATTAGTCATTTCC
TATTATCTGAAAGTTTTCTTTGTAAGTTGTAAGTATGGTCCTAAATTCAAATCCAAATTT
GATTTTGAAGATGGTACCTAATGCAGTAGTTAAATAAGTTAAAAAATGAAGGATCTAT
AATTCCTCT

>G364 Amino Acid Sequence (domain in AA coordinates: 54-76)
MDYQPNLSLRLSLPSYKNHQLNLELVLEPSSMSSSSSSSTNSSCLEQPRVFSCNYCQRK
FYSSQALGGHQNHLKRLTLAKSRELFSSNTVDSQPYPPFSGRFELYGRGYQGFLMSG
GSRDFSAARRVPESGLDQDQEKSHLDLSLRL*

>G365 (69..755)
CAATTCCTTTTACTTTCATTCTCTTTATATATTCTCTCTACGCTATAATATATATTACACA
GAATATACATGGAACCGTCCATCAAAGGAGATCAAGAAATGTTAAAAATCAAGAAACAAG
GTCATCAAGATCTTGAGTTGGGGTTGACCCCTTTTGTCACGTGGAACCGCGACCTCATCAG
AGCTCAATCTCATCGATTCTTTCAAAACCAGCTCATCATCGACTTCTCATCATCAGCACC
AGCAAGAACAATTGGCAGATCCGAGAGTGTTCTCGTGTAATTATTGTCAAAGAAAGTTCT
ATAGTTCAACAAGCGCTAGGCGGTACCAAAACGCTCATAAACGTGAGCGCACCTTAGCCA
AACGTGGACAGTATTACAAGATGACTCTCTCCTCCTTGCTTCTTCAGCGTTTGCGTTTG
GCCACGGTTCAGTCAGCAGATTCCGCAAGCATGGCATCGTTACCATTACATGGCTCGGTGA
ATAACAGGTCAACGTTAGGGATTCAAGCTCATTCAACGATCCATAAGCCAGCTTCTTAG
GAAGACAAACGACGAGTTTAAAGTCATGTTTTCAAACAGAGCATTCACCAGAAACCGACCA
TAGGAAAGATGTTGCCGAGAAATTCACCTTGAAGTCGCCGAAATAATAACAGTAACA
TGGTTGCTGCTAAGTTGGAGAGAATTGGACATTTCAAGAGCAACCAAGAAGATCATAATC
AGTTTAAAGAAAATTGACTTNGACTCTTAAGCTATGAGCTCTGCCATCTTCTTTTTAGTCTT
CATTATAACTTTTTTTATTCTCATCTTTGTTTGATATAATGATTGACGGCAGGGTGTT
AGAGTTTCACTAATGATCAAGTTGTACTTTTTATATATTTTCATTGATACCTTGTGATGT
AATTCAATATTTTAGGTCTGTTTTT

>G365 Amino Acid Sequence (domain in aa coordinates: 70-90)
MEPSIKGDQEMLIKIKQGHQDLELGLTLLSRGTATSSSELNLIDSFKTSSSSSTSHHQHQQE
QLADPRVFCSCNYCQRKFYSSQALGGHQNHLKRLTLAKRGQYKMTLSSLPSSAFAGHG
SVSRFASMASLPLHGVSNNRSTLGIQAHSTIHKPSFLGRQTTLSLHVFKQSIHQKPTIGK
MLPEKPHLEVAGNNSNMVAAKLERIGHFKSNQEDHNQFKKIDLTCLKL*

>G367 (1..708)
ATGGACGCTTCAATAGTTTTCCTCATCCACTGCTTTTCCATATCAAGATTCTCTAAACCAG
AGCATCGAAGACGAAGAAAGAGACGTTTCATAATTCTAGTCACGAACCTCAATCTCATCGAC
TGCTAGACGACACAACGAGTATCGTTAACGAATCTACAACATCCACAGAACAAGAGCTT
TTCTCATGCAACATATTGTCAAAGAACTTTCTATAGCTCACAAGCACTTGGTGGTCACCAA
AACGCACACAAGAGAGAGAACGTTGGCGAAGAGAGGACAACGTATGGCAGCGTCAGCC
TCAGCTTTTGGACATCCTTACGGTTTCTCTCCACTTCTTTCCACGGACAGTACAACAAC
CATAGGTCTTTAGGGATCCAAGCGCATTCGATAAGCCACAAGCTAAGTTCTTATAACGGG
TTTGGTGGTCATATGGTCAGATCAACTGGTCAAGACTTCCATTTGATCAACAACCAGCC
ATAGGTAAATTTCCCTCAATGGATAATTTTCATCATCATCATCATCAGATGATGATGATG
GCTCCTTCAGTAAATTCACGGTCCAATAACATCGATAGCCCAAGCAACACAGGACGGGTT
CTAGAAGGGTCACCGACTCTTGAACAATGGCACGGAGACAAAGGATTGTTGTTAAGTACA
AGTCATCATGAAGAGCAGCAGAAAATTGACTTGTCCCTCAAGCTTGA

>G367 Amino Acid Sequence (domain in AA coordinates: 63-84)
MDASIVSSSTAFPHYQDSLNSIEDEERDVHNSSELNLIDCIDDTSIVNESTTSTEQKL
FSCNYCQRTFYSSQALGGHQNHLKRLTLAKRGQRMASASAFGHPYGFSPLPFHGQYNN
HRSGLQAHSSISHLKSSYNGFGGHYQGQINWSRLPFDQQAIGKFPSPMDNFHHHHHQMOMM
APSVNSRSNNIDSPSNTGRVLEGSPTLEQWHGDKGLLLSTSHHEEQKLDLSLKL*

>G373 (1..1854)
ATGGCGATTGAAAACCTCAGCTTCTTGGCAGCGTGACGGTGTGTATGCGGTGTCAGGTG
AATCCTCCGTCAGAAGAGACTCTCACTTGTGGCAGCGTCCGTCCTCCATGGCACGTGCCG
TGTCTCCTCCCCGAATCACTCGCTTCTTCCACTGGAGAGTGGGAGTGTCCCATTGCTCC
GGCGTTGTCTGTTCCCTCCGCCGCTCCGGGTACCGGAAACGCTCGACCTGAATCTCCGGT
TCAGTTCTCGTTGCTGCGATCCGTGCGATTCAAGCTGATGAGACTTTAACCGAAGCTGAG

AAAGCCAAAAAAGGCAGAACTGATGAGTGGGGGTGGTGACGATGGTGTGATGAAGAA
GAGAAGAAGAAGTTAGAAATCTTTTGTCTATTTCGATTCAATTGCCAGAAAGACCTATC
ACGACACCGTGTGGGCACAATTTCTGTTTGAATGTTTCGAGAAATGGGCAGTAGGTCAA
GGGAAGCTAACTTGTATGATATGCCGAAGCAAAATTCGAGACATGTGGCAAAAAATCCT
CGCATCAACTTAGCTCTAGTTTCTGCTATTTCGTTTAGCAATGTTACCAATGTTCTGTT
GAGGCAACTGCAGCCAAGGTTTCATCATATTATCCGCAACCAAGACCGTCTGAGAAAGCA
TTTACTACCGAGCGGCAGTAAAACTGGGAAAGCTAATGCTGCTAGCGGTAAGTTTTTT
GTGACAATACCTCGTGATCATTTTGGTCCCATACCAGCTGAGAATGATGCTCACTAGAAAAG
CAAGGTGTTTTGGTTGGAGAATCTTGGGAGGACAGGCAAGAGTGTAGGCAGTGGGGAGCT
CATTTCCCGCATATTGCTGGCATTGCCGGCAATCAGCGGTTGGAGCTCAGTCTGTGGCC
CTCTCTGGAGGTTATGACGATGATGAGGATCATGGTGAATGGTTTTCTCTACACAGGAAGT
GGTGGGAAGGGATCTCAGTGGAAACAAAAGAATTAACAAGAAAACAGTCGCTGACCAGGCG
TTTAAAAACATGAATGAATCTCTAAGACTTAGTTGCAAAATGGGCTATCCTGTCCGAGTT
GTCAGGTCTTTGGGAAGGAGAAGCGTTCTGCATATGCCCCCTGCTGAAGGTGTGAGATATGAT
GGGGTCTATCGAATTGAGAAGTGTCTGGAGTAATGTTGGAGTACAGGGTTCTTTTAAGGTC
TGTCGTTACCTGTTTGTAGATGTGACAATGAGCCAGCTCCATGGACCAGTGATGAGCAT
GGCGATCGTCCAAGACCGTTGCCAATGTTCCGGAGCTTGAGACTGCTGCTGACCTGTTT
GTGAGAAAGGAGAGTCCATCATGGGATTTTCGATGAAGCTGAGGGTCTGTGGAAATGGATG
AAGTCTCCTCTGTTAGCAGAATGGCTTTGGATCCTGAGGAGAGGAAGAAGAATAAGAGA
GCAAAAAATACATTAAGGCCAGACTTCTGAAAGAATTTAGTTGCCAAATCTGTCCGGAA
GTGCTGAGTCTTCCAGTGACGACGCCCTTGTGCACACAACCTCTGCAAGCATGCTTAGAA
GCGAAGTTTGTCTGGGATAACTCAACTGAGAGAGAGAAGCAATGGCGGACGTAAACTACGT
GCAAGAAGAACATCATGACCTGCCCTTGCTGCACGACGGATCTCTCCGAGTTTCTCCAA
AACC CGCAGGTGAACAGAGAGATGATGGAGATAATAGAGAATTTTAAGAAGAGTGAGGAA
GAGGCTGATGCATCCATTTCTGAAGAAGAAGAAGAATCCGAACCTCCAACCTAAGAAG
ATTAAGATGGATAACAACCTCTGTTGGTGGTAGTGGTACAAGTCTCTCAGCTTAA

>G373 Amino Acid Sequence (domain in AA coordinates: 129-168)
MAIETQLPCDGDGVCMRQVNPSEETLTGCTCVTPWHVPCLLPESLASSTGEWECPCDS
GVVVPSPAAPGTGNARPESGSLVLAIRAIQADETLTEAEKAKKRQKLMSSGGDDGVDEE
EKKKLEIFCSICIQLPERPITTPCGHNFCLKCFEKWAVGQGLTCMICRSKIPRHVAKNP
RINLALVSAIRLANVTKCSVEATAAKVHHIIRNQDRPEKAFTTERAVKTKKANAASGKFF
VTIPRDHFGPIPAENDVTRKQGVLVGESWEDRQECROWGAHFPPIAGIAGQSAVGAQSV
LSGGYDDDDHGEWFLYTGSGGRDLGNKRINKKQSSDQAFKNMNESSLRLSCKMGYPVRV
VRSWKEKRSAAYAPAEGVRYDGVYRIEKCWSNVGVQGSFKVCRYLFVRCDNEPAPWTSDEH
GDRPRPLPNVPELETAADLFVRKESPSWDFDEAEGRWKMKSPVSRMALDPEERKKNKR
AKNTMKARLLKEFSCQICREVLSPVTTCAHNFKKACLEAKFAGITQLRERSNGGRKLR
AKKNIMTCCPCTTDLSEFLQNPQVNREMEIIEFKSEEEADASISEEEEESEPPTKK
IKMDNNSVGGSGTSLSA*

>G396 (1..957)
ATGGGGGAAAGAGATGATGGGTTGGGTTGAGTCTAAGCTTGGGAAATAGTCAACAAAAA
GAACCATCTCTGAGGTTGAATCTTATGCCGTTGACAACCTCTTCTTCTTCTTCTGTTT
CAACACATGCACAATCAGAATAACAATAGCCATCCCAGAAAGATTACATAACATCTCTTGG
ACTCATCTGTTTCAATCTTCTGGGATTAAACGTACAACCTGCAGAGAGAAACTCCGACGCC
GGGTCAATTCTAAGAGGTTTCAACGTGAACAGAGCTCAGTCTTCCGTGGCGGTAGTGGAC
TTGGAAGAAGAAGCCGCCGCTGCTCTCGTCTCCAAACAGCGCCGTTTCGAGTCTGAGTGG
AATAAAAGGGATCTTGGCGTGGCGAGAGGAGAGATGAAAACGAGGCGGAGAGAGCTTCT
TGCTCACGCGGAGGGGGAAGCGGTAGCGACGATGAAGACGGCGGAAACGGCGACGGA
TCAAGGAAGAACTACCGTTATCGAAGGATCAAGCTCTTGTCTCGAGGAGACTTTTAA
GAACATAGCACTCTTAATCCGAAGCAAAAGCTGGCTCTAGCAAAACAGTTGAATCTAAGG
GCAAGACAAGTTGAAGTGTGGTTTTCAGAACCGTAGGGCAAGGACGAAGCTGAAACAAACG
GAGGTTGATTGTGAGTATTTAAAGAGATGTTGCGATAATCTGACCGAGGAGAATCGACGG
CTGCAGAAAGAAGTGTCCGAGCTGAGGGCGTTGAAGTTGTCTCCACATCTCTACATGCAC
ATGACTCTCTACTACTCTCACCATGTGCCCTTCTTGGCAACGTGTCTCCTCCTCTGCC
GCCACTGTGACCGCTGCTCCTTCCACTACTACTACTCTACGGTGGTGGGGCGGCCAAGT
CCACAGCGATTAACTCCTTGGACTGCTATTTCTCTCCAGCAAAAATCAGGTGCTAG

>G396 Amino Acid Sequence (domain in AA coordinates: 159-220)
MGERDDGLGLSLSLGNSQQKEPSLRNLMLPLTSSSSSFQHMHNQNNNSHPQKIHNISW

[illegible]

>G479 (1..1128)

ATGGAGATGGGGTTCCAACTCGGGTCCGGGTGTCATGGTCCGGGTGAGGCAGAGTCCGGTGGT
TCCTCCACTGAGTGCATCTCTTTTCAAGTGGAGGGGCTCATGTTTGCCAGAGAAGATCTACTTC
GAGGACGGTGGTGGTGGATCCGGGTCTTCTTCCCTCAGGTGGTCGTTCAAACAGACGTGTC
CGTGAGGCGGGGTCCGGTTCAGTCCGGGTGAGATACCAAGGTGCCAAGTGGGAAGGTTGTGGG
ATGGATCTAACCAATGACAAAGGTTATTAATCGAGACACCGAGTTTGTGGAGTGCACTCT
AAAAACCTTAAAGTCACTGTGGCTGTATCTCGAAGACAGAGGTTTGTGCAACAGTGCAGCAGG
TTTCATCAGCTTCCGGAAATTTGACCTAGAGAAAAGGAGTTGCCGCAGGAGACTCGCTGGT
CATAATGAGCGACGAAGGAAGCCACAGCCTGCGTCTCTCTCTGTGTTAGCTTCTCGTTAC
GGGAGGATCGCACCTTTCGCTTTACGAAATGGTGATGCTGGAAATGAATGGAAGCTTTCTT
GGGAACCAAGAGATAGGATGGCCAAGTTCAAGAACATTGGATACAAGAGTGATGAGGCGG
CCAGTGTCTGTCACCGTCAATGAGAGCATCAATCCAATGAATGTATTAGTCAAGGTTCAGTT
GGTGGAGAGGGGACAAGCTTCTCATCTCCAGAGATTATGGACATAAACTAGAGAGCTAC
AAGGGAAATTGGCGACTCAAACCTGTGCTCTCTCTCTCTGTGCAATCCACATCAACCACAT
GACAACAACAACAACAACAACAACAGCAACAACAACAACAATACATGGCGAGCTTCT
TCAGGTTTTGGCCCGATGACGGTTACAATGGCTCAACCACCACCTGCACCTAGCCAGCAT
CAGTATCTGAACCCGCCCTTGGGTATTCAGGACAATGATAATGATATGTCTCCTGTTTTG
AATTTAGGTCGATACACCGAGCCGAGATAATTTGTCTAGATAAGTAGTGCCACGGCAATGGGT
GAGTTTCGAGTTATCTGATCACCATCATCAAAGTAGAGACAGTACAGTATGGAAGATGAGAAC
ACAAAGGGCTTATGACTCTTCTTCTCACCATCAACCACTGGTCTCTCTGA

129

FHQLPEFDLEKRSRRRLAGHNERRRKPQPASLSVLASRYGRIAPSLYENG DAGMNGSFL
GNQEIGWPSSRTLDTRVMRRPVSSPSWQINPMNVFSQGSVGGGTSFSSPEIMDTKLESY
KGIGDSNCALSLLSNPHQPHDNNNNNNNNNNNTWRASSGFGPMTVTMAQPPAPPSQH
QYLNPPWVFKDNDNDMSPVNLGRYTEPDNCQISSGTAMGEFELSDHHHQSRRQYMEDEN
TRAYDSSSHHTNWSL*

>G546 (1..588)

atgactcgaccgtcaagattacttgagacggcgccgaccaccacacacgcgcgaggag
atgatcgacgcgaatccgacatgggtggatcttgctcggtctcttcttgctctctatc
tgcttgctggtctcgccgcgctcgatcgatgcgcttggtccggcggtttacagccgga
ggagattcgccgtcaccgaacaaaggcttgaaaaagaaagctcttcagtctcttccaaga
tccactttcaccgcgcggaatcaacctccggcgccgcgctgaagaggagactcgacg
gaatgtgctatttgctcactgacttcgcccgcggtgaagaaataagagtgtctctctt
tgtgtgctattcttccacgtggagtgtattgacaaatggctagtttctaggtctcttgt
cttcttgctcgaggattcttaccgcggtgagatgtgaccggtgtggtcatgcttctacg
gcggagatgaagatcaagctcatcgatcatcaacatcaccaacactcttctactaccatt
cctacgtttcttctcttaa

>G546 Amino Acid Sequence (domain in AA coordinates:114-155)

MTRPSRLLETAAPPPQPSEEMIAAESDMVILSALLCALICVAGLAADVRCALRRFTAG
GDSPPNKGKLLKALQSLPRSTFTAAESTSGAAEEGDSTECALCLTDFADGEEIRVLPL
CGHSFHVCEIDKWLVSRSRSCPSRRILTPVRCDRCGHASTAEMKDQHRHQHQSSTTI
PTFLP*

>G551 (1..708)

ATGGAGTGGTCAACAACGAGCAACGTAGAAAACGTGAGAGTAGCTTTTCATGCCACCGCCA
TGGCCGGAGTCTAGTTCCTTTAACTCGCTCCACAGCTTCAACTTTGATCCTTACGCAGGA
AATTCATATACGCTTGGCGATACACAAACCGGACCGGTTATCTCTGTACCGGAATCAGAA
AAGATCATGAATGCGTACCGATTTCGGAACAACAACATGAGATGATAAAAAAGAAGAGA
CTAACGAGTGGACAATTAGCTTTCATTTGAGCGAAGTTTCAAGAAGAGATCAAATTAGAT
TCAGACAGGAAGGTGAAGCTGTGAGAGAGCTCGGTCTGCAGCCACGTCAGATAGCAGTT
TGGTTCCAAAACCGCGTGCACGGTGAAGGCGAAGCAGCTTGAGCAGTTGTACGACTCG
CTTAGACAAGAGTACGACGCTCGTTTCTAGGGAGAAACAAATGTTACACGATGAGGTGAAG
AAGCTGAGAGCTTTACTAAGAGACCAGGGTTTGATCAAGAAGCAAAATCTCTGCCGGGACC
ATCAAAGTTTCCGGTGAGGAAGACACGGTGGAGATTTTCATCGGTGGTGGTAGCTCATCCA
AGAACGGAGAATATGAACGCAAAATCAAATCACCGGAGGGAATCAAGTTTACGGTCAATAC
AACAAATCCGATGCTGGTTGCTTCTCTGCTGGCGGTCATACCCCTGA

>G551 Amino Acid Sequence (conserved domain in AA coordinates:73-133)

MEWSTTSNVENVRAFMPPWPPESSFNLSHFNFDPYAGNSYTPGDTQTGPVISVPESE
KIMNAYRFPNNNNEMIKKRLTSGQLASLERSFQEEIKLSDRKVKLSRELGLQPRQIAV
WFQNRARRWAKQLEQLYDSLROEYDVVSREKQMLHDEVKKLRALLRDQGLIKKQISAGT
IKVSGEEDTVEISSVVVAHPRTENMNANQITGGNQVYGQYNNPMLVASSGWPSYP*

>G578 (1..978)

ATGCATAGTTTGAATGAAACAGTAATTCCTGATGTTGATTACATGCAGTCTGATAGAGGG
CATATGCATGCTGCTGCCTCTGATTCCAGTGATCGATCAAAGGATAAGTTGGATCAAAAG
ACCCCTTCGTAGGCTTGCTCAAATCGTGAGGCAGCAAGAAAAAGCAGATTGAGGAAGAAG
GCGTATGTTACGACGCTGGAAGATAGTCGATTAAAGCTGACTCAAGTTGAGCAGGAGCTG
CAAAGAGCAAGACAGCAGGGAGTTTTCATCTCAAGTTCAGGAGACCAAGCTCATTCTACT
GGTGGCAATGGTGGGGCTTTGGCATTGTGATGCAGAACTCACGATGGCTTGAAGAAAAG
AACAGGCAAAATGAACGAGCTGAGATCTGCCCTGAATGCTCATGCAGGTGATACTGAGCTC
CGGATAATTGTGGATGGAGTGATGGCTCACTATGAGGAGCTTTTCAGGATTAAAGAGCAAT
GCATCTAAGAAATGATGTCTTCCACTTGTATCTGGAATGTGGAAAACACAGCTGAGCGA
TGTTTCTTGTGGCTTGGCGGGTTCCCGTCATCCGAACCTTCTCAAGCTTCTTGGCAATCAG
CTAGAGCCCATGACAGAACGACAGGTAATGGGCATCAATAGCTTGCAGCAGAGCTCGCAG
CAGGCAGAGATGCTTTATCTCAAGGGATGGAGAGTTTACAGCAATCCCTAGCTGATACT
TTATCCAGTGAACCTCTTGGTTCCAGTTCATCGGATAATGTGCGGAGCTACATGGGTCAG
ATGGCCATGGCAATGGGCAAGTTAGGCACCCTCGAAGGATTCATACGCCAGGCTGATAAC
TTGAGGCTGCAAACTACAAACAGATGCTTCGAGTATTAACAACACGTCAGTCAGCTCGT
GCTCTTCTTGTATACAGGATTATTCATCTCGATTACGTGCTCTTAGTTCCTTGTGGCTT
GCCCGGCAAGAGAGTGA

>G578 Amino Acid Sequence (domain in AA coordinates 36-96)
MHSLNETVIPVDVYMQSDRGHMHAAASDSSDRSKDKLDQKTLRRLAQNREAARKSRLRKK
AYVQQLSRLKLTQVEQELQRRARQQGVFISSSGDQAHSTGGNGGALAFDAEHSRWLEEK
NRQMNELRSALNAHAGDTELRIIVDGVMAHYEELFRIKSNASKNDVFHLLSGMWKTPAER
CFLWLGGFPSSSELLKLLANQLEPMTERQVMGINSLQQTSSQAEDALSQGMESLQQSLADT
LSSGTLGSSSSDNVASYMGQMAMAMGKLTLEGFIRQADNLRQLQLQMLRVLTTTRQSAR
ALLAIHDYSSRLRALSSLWLARPRE*

>G596 (168..1121)

TAATTTCTCTACTTCAGATTTTTTCTCCTTAGATTAATTTAATTGAGTTATTGTACATC
CCTCAAGCTAAGATTCTGTTTTGTGAGTTGAGTGGATGAGAAGAGGAGAGATTAAGTAA
ATTAGGGTTCAATTGTTTACTTTTTGTGTTTTTATATCAAGTAATGGATCAGGTCT
CTCGCTCTCTCTCCACCTTTTCTCTCAAGAGATCTCCATCTTCACCCACACCATCAAT
TCCAGCATCAGCAGCAGCAGCAACAGAATCACGGCCACGATATAGACCAGCACCGAA
TCGGTGGGCTAAAACGTGACCGAGATGCTGATATCGATCCCAACGAGCACTCTTCAGCCG
GAAAAGATCAAAGTACTCCTGGCTCCGGTGGAGAAAGCGGCGGAGGAGGAGGAGATA
ATCACATCAGCAGAGGCCACGTGGCAGACCAGCGGGATCTAAGAACAAACCAACCGC
CAATCATCATCTACTCGAGCAGCGCAACCGCTCTCAATCTCATGTTCATGGAAGTAGCAA
ACGGATGTGACGTATGGAAGTGTACCGTCTTCGCTCGCCGTCGCCAACGTGGCATCT
GCGTTTTGAGCGGAAACGCGCCGTTACCAACGTTACCATAAGACAACAGCTTCAGTAC
CTGGTGGTGGCTCATCTGTCGTTAACTTACACGACGTTTCGAGATTCTTCTCTCTCGG
GATCATTCTCTCTCTCCGCTCCACAGCTGCGTCAGGTCTAACGATTTACTTAGCCG
GTGGTCAGGGACAGGTTGTTGGAGGAAGCGTGGTGGTCCACTCATGGCTTCAGGACCTG
TAGTGATTATGGCAGCTTCGTTTGGAAACGCTGCGTATGAGAGACTGCCGTTGGAGGAAG
ACGATCAAGAAGAGCAAAACAGCTGGAGCGGTTGCTAATAATATCGATGGAAACGCAACAA
TGGGTGGTGGAAACGCAACGCAAACTCAGACGACGAGCAACAGCAACACAGTTGATGC
AAGATCCGACGTCGTTTATACAAGGGTTCCTCCGAATCTTATGAATTCTGTTCAATTGC
CAGCTGAAGCTTATTGGGGAACCTCCGAGACCATCTTTCTAAATCGCGAAGAAAAACAAG
TTAGATACGTTTCGTTGTTTTTAATTTATAATCTCTCTTCTGTCAAGTTTAAATTTCTTT
TTCTTCTCTTTGTTTTCTAAAGATAATTGTAGTCTTTGACGAAGATTCGTGGTACGTAT
GAATCGAAGAGAATCGTTTTGTCATGGGATTGCTCGATCTATTAGGTTTGAGAGGGGGT
TTGTGTTTTGCGTTGACTAGCAGATTATAAAATTGTTGATTTTCGAGTTTATTATTTTCAT
GTGTTGGTGATAAA

>G596 Amino Acid Sequence (domain in AA coordinates: 89-96)

MDQVSRSLPPPLSRDLHLHPHHQFQHQQQQQNHGHDIDQHRIGGLKRDADIDPNE
HSSAGKDQSTPGSGGESGGGGGDNHITRRPRGRPAGSKNPKPPIIITRDSANALKSHV
MEVANGCDVMESVTVFARRRQRGICVLSGNGAVTNVTIRQPASVPGGSSSVNHLHGRFEI
LSLSGSFLPPPAPPAASGLTIYLAGGQGVVGGSVVGPLMASGPVVIMAAASFGNAAYERL
PLEEDDQEEQTAGAVANNIDGNATMGGGTQTQTQTQQQQQQLMQDPTSFIQGLPPNLMN
SVQLPAEAYWGTTPRPSF*

>G617 (59..1141)

CAGATCTGTTCTTTACACCAAATTGAGTACTGAAGATCTTGTGAGTGAATTAAAGAGAT
GAGATCAGGAGAAATGTGATGAAGAGGAGATTCAAGCAAAGCAAGAAAGAGATCAAAATCA
AAATCATCAAGTAAACTTAAACCACATGTTGCAACAACAACAGCCGAGTTCCGTATCATC
TTCAAGGCAATGGACTTCAGCTTTTAGGAATCCAAGAATCGTTCGAGTCTCAAGAACATT
CGGTGGCAAAGACAGACACAGCAAAGTATGTACAGTCCGTGGTCTTCGAGACCGGAGGAT
AAGGTTGTCCGTACCTACAGCTATTCAACTCTACGACCTCAAGATCGATTAGGGCTGAG
TCAGCCAAGCAAAGTCATTGATTGGTTACTCGAAGCAGCAAAGATGACGTAGACAAGCT
ACCTCCTCTACAATCCACATGGATTTAACCAGATGTATCAAATCTCATCTTCGGAAA
CTCCGGGTTTGGAGAATCTCCATCTTCAACTACATCAACAACGTTTCCAGGAACCAATCT
CGGGTCTTGGAAAATTGGGATCTTGGTGGTTCTTCAAGAACAAGAGCAAGATTAACCGA
TACAACCTACGACCCAAAGAGAAAGTTTTGATCTTGATAAAGGAAAATGGATCAAAAACGA
CGAGAATAGTAATCAAGATCATCAAGGGTTTAAACACCAATCATCAACAACAAATTCCTCT
GACCAATCCGTACAACAACACTTCAGCTTATTACAACCTTGGACATCTTCAACAATCGTT
AGACCAATCTGGTAATAACGTTACTGTGCAATATCTAATGTTGCTGCTAATAATAACAA
TAATCTCAATTTGCATCCTCTCTCCTCGTCTGCCGGAGATGGATCTCAGCTTTTTTTCGG
TCCTACTCCTCCGCAATGAGCTCTCTATTCCCGACATACCTTCGTTTCTTGGAGCTTC
TCATCATCATCATGTCGTGATGGAGCCGGTCATCTTCAGCTCTTTAGCTCGAATTCAAA

TACCGCATCGCAGCAACACATGATGCCGGGTAATACGAGTTTGATTAGACCATTTCATCA
TTTGATGAGCTCGAATCATGATACGGATCATCATAGTAGCGATAATGAATCAGATTCTTG
AATGATTTTATATATCTACACTATACATTGAAAAATGTTATATGTATACGTATTCTTCTAT
ATTTTGATATATATGCGTATTGTTGGATTGGTTTATGTATCT
>G617 Amino Acid Sequence (domain in AA coordinates: 64-118)
MRSGECDEEEIQAKQERDQNNHVNLMHMLQQQPPSSVSSSRQWTSAFRNPRIVRVSRT
FGGKDRHSKVCTVRGLRDRIRLSVPTAIQLYDLQDRLGLSQPSKVIDWLLEAAKDDVDK
LPPLQFPHGFMQMYPNLIFGNSGFGESPSSTTSTTFPGTNLGFLENWDLGGSSRTRARLT
DTTTTQRESFDLDKGKWKINDENSNDHQGFNTNHQQQFPLTNPYNNTSAYYNLGHLLQS
LDQSGNNVTVAISNVAANNNNNLNLHPPSSSAGDGSQLEFFGPTPPAMSSLFPTYPSPFLGA
SHHHHVVDGAGHLQLFSSNSNTASQQHMMPGNTSLIRPFHMLMSSNHDTHHSSDNESDS
*
>G620 (40..666)
GAATTGAACCTGGACCAGCACAGCAACAACCCAACCCAATGACCAGCTCAGTCATAGTA
GCCGGCGCCGGTGACAAGAACAATGGTATCGTGGTCCAGCAGCAACCACCATGTGTGGCT
CGTGAGCAAGACCAATACATGCCAATCGCAAACGTGATAAGAATCATGCGTAAACCTTA
CCGTCTCACGCCAAAATCTCTGACGACGCCAAAGAAACGATTCAAGAATGTGTCTCCGAG
TACATCAGCTTCGTGACCGGTGAAGCCAACGAGCGTTGCCAACGTGAGCAACGTAAGACC
ATAACTGCTGAAGATATCCTTTGGGCTATGAGCAAGCTTGGGTTGATAACTACGTGGAC
CCCCTCACCGTGTTTCATTAAACCGGTACCGTGAGATAGAGACCGATCGTGGTTCTGCACTT
AGAGGTGAGCCACCGTCGTTGAGACAAACCTATGGAGGAAATGGTATTGGGTTTCACGGC
CCATCTCATGGCTACCTCCTCCGGGTCCTTATGGTTATGGTATGTTGGACCAATCCATG
GTTATGGGAGGTGGTTCGTTACTACCAAACCGGTCGTCGGGTCAAGATGAATCCAGTGTT
GGTGGTGGCTCTTCGTTCTCCATTAAACGGAATGCCGGCTTTTGACCATTATGGTCAGTAT
AAGTGAAGAAGGAGTTATCTTCATTTTATATCTATTCAAAACATGTGTTTCGATAGAT
ATTTTATTTTATGTCTTATCAATAACATTTCTATATAATGTTGCTTCTTTAAGGAAAAG
TGTTGTATGTCAATACTTTATGAGAACTGATTATATATGCAAT
>G620 Amino Acid Sequence (domain in AA coordinates: 20-118)
MTSSVIVAGAGDKNGIVVQQPPCVAREQDQYMPIANVIRIMRKTLP SHAKISDDAKET
IQECVSEYISFVTGEANERCQREQRKTITAEDILWAMSKLGFNDYVDPLTVFINRYREIE
TDRGSALRGEPPSLRQTYGGNGIGFHPGSHGLPPPGPYGYGMLDQSMVMGGGRYYQNGSS
GQDESSVGGGSSSSINGMPAFDHYQYK*
>G625 (151..1137)
AATCGACCATTCACAACGATGACATTCAAACACTCTTCAGTTTCCCTTCTCTTCTGATTC
GTCCTCTCCACTATTTTCTCAATTTCTTTAATCTCTCTTTCTCTCTACTTCTCTCT
TCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AACCATCTGGAAGATAATAACCAAACCTAACCATAATAATCCTCAATCCGATTCCACC
ACCGACTCATCACTTCTCCGCTCAACGCAAACGCAAAGGCAAAGGTGGTCCGGACAAC
TCCAAGTTCCGTTACCGTGGCGTTCGACAAAGAAGCTGGGGCAAATGGGTCCGCCGAGATC
CGAGAGCCACGTAAGCGCACTCGCAAGTGGCTTGGTACTTTCGCAACCGCCGAAGACGCC
GCACGTGCCCTACGACCGGGCTGCCGTTTACCTATACGGGTACGTGCTCAGCTCAACTTA
ACCCCTTCGCTCTCCTTCTCCTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
CCTTCCACCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
GCCACCGTAGGAGGAGGACCAACTTTGGTCCGTACGGTATCCCTTTTAAACAACAACATC
TTCCTTAATGGTGGGACCTCTATGTTATGCCCTAGTTATGGTTTTTTCCTCAACAACAA
CAACAACAAAATCAGATGGTCCAGATGGGACAATTCCAACACCAACAGTATCAGAATCTT
CATTCTAATACTAACAATAACAAGATTTCTGACATCGAGCTCACTGATGTTCCGGTAACT
AATTCGACTTCGTTTCATCATGAGGTGGCGTTAGGGCAGGAACAAGGAGGAAGTGGGTGT
AATAATAATAGTTCGATGGAGGATTGAACCTCTAGCTGGTTCGGTGGGTTCGAGTCTA
TCAATAACTCATCCACCGCGTTGGTTGATCCGGTATGTTCTATGGGTCTGGATCCGGGT
TATATGGTTGGAGATGGATCTTCGACCATTGCGCTTTTGGAGGAGAAGAATAATAGT
CATAATTGGGGGAGTATTTGGGATTTTATTGATCCCATCTGGGGGAATTTCTATTAAATT
GTTTTTGTGGAAGATCATATTATATACGATGAGCATCCCTAAGGTCGGTCAAGAGCATTG
GAGATTCAATTGTTGAGAGGAATCAAAGAGATTGCATTCTATGAGGAGCTCTGCATGCAAA
ATTTTGGAGGATTTTCTTACTACCTATAGAGATAAATAAGAGGATTTTATTATTATT
TTGAAGATTTTATTCTCAAGGAATTCGTAAAAGAGATTACGGTTCCAATAAAGTATGTA
TATGTGAAGAGAATCGGAGGAGATGGTGGAAAGTTGTATGGGAATTTTATTGGTTCAAC

ACTCCTCCTCCTCCAACGAATTGTAGCTATAGGTTGTTGGATTGATCTCACAAGCAAT
TCTCCTGCTCAATCCCTCAAGACAAGCAACCGATGGATACTTGTGGAGCTGCCAAGTGT
CAAGAACCCATCACTCCAACCTCAATGAGTGAGCAGAAGAAGCAACAAACATCAAGAAGT
CGAATAAAGTGCAAATGCAAGGCATTGCGGTTGGTCGTGCGGTTGATTTAACTGTTG
AAATCTTACGATGAAGTGAATGATGAGCTTGAGGAGATGTTTGAAGATTCAAGGACAGCTT
CTTGCCCGAGACAAATGGATCGTTGTCTTCACTGATGATGAAGGAGATATGATGCTTGCT
GGTGATGATCCGTGGAATGAGTTTGCAGATGGCAAAGAAGATATTTATATATTTCGAGC
GATGAGGTTAAGAAATGACAACGAACTGAAGATTTCTTCGTGTTAGAGAATGAGGAA
TATGGTAATGAATCATTGCAAAATCGTAGTAGGGGTGAGAGTTTTAGCTGTTAATTAAG
GTTAATTCGCGCAGCTCGTTTTAGTGCGTAAGTGCTAAAGACTTTTTTTTAGTCTGTG
TATATAAAGTCTTGCTCTTTTTTCATGTCAATTTTCAAGTTGGCGATTAAATATTTTCG
GTTTTGGGACAGTGGTTGATGGGGCGGTTTTACATTTTTTATGTGTATGTAATTTGTTCCA
AAACCATTCATTTTCAAA

>G716 Amino Acid Sequence (domain in AA coordinates: 24-355)
MASVEGDDDFGSSSSRSYQDQLYTELWKVCAGPLVEVPRAQERVVFYFPQGHMEQLVASTN
QGINSEIPVFDLPPKILCRVLDVTLKAEHETDEVYAQITLQPEEDQSEPTSLDPPPIVGP
TKQEPHSFVKILTASDTSTHGGFSVLRKHATECLPSLDMTQATPTQELVTRDLHGFEWRF
KHIFRQPRRHLLTTGWSTFVSSKRLVAGDAFVFLRGENDLRVGVRLARHQSTMPTSV
ISSQSMHLGLVLTASHAVRTTTFIVVFYKPRISQFIVGVNXYMEAIKHGFSLGTRFRMRF
EGEESPERIFTGTIVGSGDLSSQWPASKWRSLOVQWDEPTTVQRPDKVSPWEIEPFLATS
PISTPAQQPQSKCKRSRPIEPSVKTPAPPSFLYSLPQSQDSINASLKLFDPSLERISGG
YSSNNSFKPETPPPTNCSYRLFGLDLSNSPAPIPDQKQPMDCGAACQEPITPTSMS
EQKKQOTSRSRTKVQMGIAGRAVDLTLLKSYDELIDELEEMFEIQGOLLARDKWIVVF
TDDEGDMMLAGDDPWNEFCMAKKIFIYSSDEVKMTTKLKISSLENEEYGNESFENRS
RG*

>G725 (46..1122)
CCTCTTTTCAGAGAGAGAAAGAGAGTCAGAGAGAGAGAGAGAGAGAAATGTTCCATGCTAAG
AAACCTTCAAGTATGAATGGTTTCATATGAGAACAGAGCTATGTGCGTTCAAGGCGATTCA
GGCCTTGTCTCACCACCGACCTAAACCGCGTTTGGCGTTGGACCGTCGAACCTCCACGAG
CGTTTTGTGACGCGCTCGCTCAGCTCGGCGGCCCGACAAAGCGACCCCAAGACGATT
ATGAGAGTTATGGGTGTGAAGGCTTACTCTTTACCACCTAAAGAGCCATCTTCAGAAA
TTCAGGCTTGGAAGCAGCCGACAGAGAGTACGAGATCACTCCACAAAGGAAGGTTCA
AGAGCTTCTGCCATGGATATTAGCGCAACGTAGCTTCTTCTTCTGCGATGATGAGTCGC
AACATGAATGAGATGCAAATGGAAGTGCAGAGAAGGTTGCATGAACAGCTAGAGGTGCAA
AGACATCTGCAACTGAGGATTGAAGCACAAGGAAAGTACATGCAATCTATCTTGGAGAGA
GCTTGCCAAACCTAGCCGGTGAGAACATGGCAGCCGCCACCGCAGCAGCCCGCTCGGA
GGAGGATACAAGGTAATCTGGGAAGTTCGAGTCTTTCAGCAGCGGTGGGCCACCTCCT
CATCCTCTTAGTTTTCCCGCGTTTCAAGACCTAAACATCTATGGAACACAACCGACCAA
GTCTTCGACCATCACAATTCATCATCAAAACATAGAGAACCATTTACGGGTAACAAT
GCTGCAGACACCAACATTTACTTGGGGAAGAAGCGACCTAATCCTAATTTTGGTAACGAT
GTAAGGAAAGGACTATTGATGTGGTCTGATCAAGATCAGATCTTTCGCAAACCAATCG
ATCGATGATGAGCATAGAATTCAGATACAGATGGCTACACATGTCTCCACGGATTGGAT
TCTTTGTGCGAGATCTACGAAAGGAATCAGGTTTATCAGGTGATGAAGGGAATAATGGT
GGGAAATTACTGGAAAGGCCATCGCCTAGGAGATCACCATTGAGTCTATGATGAACCTT
AATGGTGGATTAATACAAGGAAGAACTCGCCATTTGGGTGATACAATTTATTAATTTT
ATCTATGAGTGATGCATGGGAATGTAAGAACGAGATATATATGTTTGTGCTATTGTGAGTT
TGACGTAGGTTTAGAGAAAA

>G725 Amino Acid Sequence (domain in AA coordinates: 39-87)
MFHAKKPSSMNGSYENRAMCVQDGLVLTDPKPRLRWTVELHERFVDAVAQLGGPDKA
TPKTIMRVMGVKGLTLYHLKSHLQKFLGKQPHKEYGDHSTKEGSRASAMDIQRNVASSS
GMMSRNMNMQMEVQRRLEHQLVQRHLQLRIEAQKYMQSILERACQTLAGENMAAATA
AAAVGGGYKGNLSSSLAAVGGPPHPLSFPPFQDLNIYGNITDQVLDHNFHHQNIENH
FTGNNAADTNIYLGKKRPNPNFGNDVRKGLLMWSDQDHDLSANQSIDDEHRIQIQMATHV
STDLDLSLEIYERKSLSGDEGNNGKLLERPSPRRSPLSPMMPNGGLIQGRNSPFG*
>G727 (43..1977)
CTTCTTCTCTTCTCTGATCGTTTCGTTTTCTGGACGAGAGAGATGGTAAATCCGGGTCAC
GGAAGAGGACCGATTCCGGGTACTGCTGCTGGTGGGTCAAACCTCCGACCCGTTTCTGCG

AATCTTCGAGTTCTTGTCGTTGATGATGATCCAACCTTGCTCATGATCTTAGAGAGGATG
CTTATGACTTGTCTCTACAGAGAGCAGAGAGCGCATTGTCTCTGCTTCGGAAGAACAAG
AATGGTTTTGATATTGTCTATTAGTGATGTTTCATATGCCTGACATGGATGGTTTTCAAGCTC
CTTGAACACGTTGGTTTAGAGATGGATTTACCTGTTATCAATCTGAATGTTTTGAAACCT
TTGGTTATAGTGATGCTGCGGATGATTGGAAGAGCGTTGTGTTGAAAGGAGTGACTCAC
GGTGCAGTTGATTACCTCATCAAACCGGTACGTATTGAGGCTTTGAAGAATATATGGCAA
CATGTGGTGCGGAAGAAGCGTAACGAGTGGAATGTTTCTGAACATCTGAGGGAAGTATT
GAAGATACTGGCGGTGACAGGGACAGGCAGCAGCAGCATAGGGAGGATGCTGATAACAAC
TCGCTCTCAGTTAATGAAGGGAACGGGAGGAGCTCGAGGAAGCGGAAGGAAGAGGAAGTA
GATGATCAAGGGGATGATAAGGAAGACTCATCGAGTTTAAAGAAACCACGCTGGTTTGG
TCTGTTGAATTGCATCAGCAGTTTGTGCTGCTGTGAATCAGCTAGGCGTTGACAGTGAG
TTAAAAACTTGCTTGCTTATGCAATTTGTGTGTGCTGATTGGTAACATTGTGGAATCCAG
AAGTATCGGATATATCTGAGACGGCTTGAGAGGAGTATCGCAACACCAAGGAAATATGAAC
CATTGCTTTATGACTGGTCAAGATCAGAGTTTTGGACCTCTTCTTCGTTGAATGGATT
GATCTTCAATCTTTAGCTGTTACTGGTCAGCTCCCTCCTCAGAGCCTTGACACAGCTTCAA
GCAGCTGGTCTTGGCCGGCCTACACTCGCTAAACAGGGATGTCGGTTTCTCCCTTGTA
GATCAGAGAAGCATCTTCAACTTTGAAAACCCAAAAATAAGATTGGAGACGGACATGGT
CAGACGATGAACAATGGAAATTTGCTTCATGGTGTCCCAACGGGTAGTCACATGCGTCTG
CGTCTCGGACAGAATGTTTCAGAGCAGCGGAATGATGTTGCCAGTAGCAGACCAGCTACCT
CGAGGAGGACCATCGATGCTACCATCCCTCGGGCAACAGCCGATATTGTCAAGCAGCGTT
TCAAGAAGAAGCGATCTCACTGGTGCCTGGCGGTTAGAAACAGTATCCCCGAGACCAAC
AGCAGAGTGTACCAACTACTCACTCGGTCTTCAATAACTTCCCCGCGGATCTACCTCGC
AGCAGCTTCCCGTTGGCAAGTGCCCCAGGGATTTAGTTCCAGTATCAGTTTCTTACCAA
GAAGAGGTCAACAGCTCGGATGCAAAAGGAGGTTTCATCAGCTGCTACTGCTGGATTGGT
AACCAAGCTACGACATATTTAACGATTTTCCGAGCACCACAGCACAACAAGAATC
AGCAATAAACTAAACGATTGGGATCTGCGGAATATGGGATTGGTCTTCACTTCCAATCAG
GACGACGCACTGCAACCGCAACCGCAGCATTTTCCACTTCGGAAGCATACTCTTCGTCT
TCTACGCAGAGAAAAAGACGGGAAACGGACGCAACAGTTGTGGGTGAGCATGGGCAGAAC
CTGCACTACCGAGCCGGAATCTGTATCATCTGAACACGTTTTTATGACGGTGGTTCA
GTCAGAGTGAAGTCAGAAAGAGTGGCGGAGACAGTGACTTGTCTCCAGCAAATACATTG
TTTACAGAGCAGTATAATCAAGAAGATCTGATGAGCGCATTTCTCAAACAGGTTTGATTA
TTACTCGAATACAGTGCCTCTAAAC

>G727 Amino Acid Sequence (domain in AA coordinates: 226-269)

MVNPGRHGRPDGTAAGGSNSDPFANLRLVVDVDDPTCLMILERMLMTCLYREQRAHCL
CFGRTKNGFDIVISDVHMPDMDGFKLLEHVGLEMDLPVNLNLKPLVIVMSADDSKSVV
LKGVTHGAVDYLKPVRIEALKNIWQHVVVRKKRNEWNVSEHSGSIEDTGDDRDRQQQHR
EDADNNSSSVNEGNRRSSRKREEVEVDQDDKEDSSSLKKPRVVWSVELHQQFVAAVNQ
LGVDSELKTCCLMLHLCVSGNIVEFQKYRIYLRRLGGVSQHQGNMNHSMFTGQDQSFGL
SSLNGFDLQSLAVTGQLPPQSLAQLQAAGLGRPTLAKPGMSVSPPLVDQRSIFNFENPKIR
FGDGHGQTMNNGNLLHGVPTGSHMLRPGQNVQSSGMLLPVADQLPRGGPSMLPSLGQQP
ILSSSVRRSDLTGALAVRNSIPETNSRVLPTHSVFNNFPADLPRSSFPLASAPGISVP
VSVSYQEEVNSSDAKGGSSAATAGFGNPSYDIFNDFPQHQQHNKNI SNKLNWDWDLRNMGL
VFSSNQDAATATATAAFSTSEAYSSSSTQKRRETDATVVGEHGQNLQSPSRNLYHLNHV
FMDGGSVRVKSERVAETVTCPPANTLFHEQYNQEDLMSAFILKQV*

>G740 (25..924)

CTTCTTCACTTTTTTTTAAACGATGGCTTCAGAGGATCAATCGGCGGCGAGATCTACC
GGGAAGGTGAACCTGGTTCAACGCTTCTAAAGGCTATGGTTTCATTACTCTGACGATGGC
AGCGTAGAGCTTTTGTTCATCAATCTTCAATTTGTCTCCGAAGGTTACCGGAGTTTAAAC
GTCGCGGATGCGGTTGAGTTGCTATTACTCAGGAAGCGACGGTAAGACTAAAGCCGTC
AATGTTACTGCTCCTGGTGGTGGTTCTCTCAAGAAGGAGATAACTCTCGTGGTAACGGT
GCTAGCGCGCGCGGCGGTGGAAGCGGTTGCTACAATTGCGGTGAGTTAGGTTCATATCTCT
AAAGATTGTGGTATTGGTGGCGGCGGCGGAGGTGGTGAACGTAGATCTAGAGGAGGAGAA
GGTTGTTACAATTGTGGTGATACCTGCTCACTTCGCTAGGATGTACTTCAGCTGGAAAC
GGTGACCAACCTGGAGCCACCAGGTGGAAACGATGGTTGCTACACTTGCCTGATGTT
GGTCAGTGGCTAGGGATTGTACTCAGAAATCAGTTGGAACCGGAGACCAACGTGGAGCG
GTCAAAGGTGGAAACGATGGTTGCTACACTTGTGGTGTGTTGGTCACTTTGCTAGGGAT
TGTAATCAGAAGGTTGCTGCCGGAACGTCAGAACCGGTGGTGGTGGTAGTGAACCTTGT

TATTCATGCGGTGGAGTTGGTCACATTGCAAGAGATTGTGCGACTAAGAGACAGCCTTCT
CGTGGGTGTTACCAGTGTGGTGGTCTGGTCACTTGGCTCGTGATTGTGACCAGAGAGGA
AGCGGTGGAGGAGGTAATGATAATGCGTGCTACAAGTGTGGTAAGGAAGGTCACCTTTGCA
AGGGAATGTTCTTCTGTAGCTTAATCGATTTCCTAATCAACAAAACAAAAACAAGAAT
GAAATTGAATCGAGTTATATAGTTTGGTATATATACTCTTCGTTTTTCATTTATCTTTTT
TTTTGTTGTTGATGGGAATGAAATTGCCTGGTCCCTTTTGGTGTGTTTTTGAGCTTTTATT
ATTATACAGAGTGTATCCCTTTTTTGTATAACTATTACAAGTTTTTAGCTTTATTTGATA
TGGATGCTCTCTCTTTTCTTCTATCTGTTTCTGGAAATTTTGACCTCATCATATTACTT
ATGTCATCCAAA

>G740 Amino Acid Sequence (domain in AA coordinates: 24-42, 232-268)
MASEDQSAARSTGKVNWFNASKGYGFITPDDGSVELFVHQSSIVSEGYRSLTVGDAVEFA
ITQSGDGKTKAVNVTPAGGSLKKENNSRNGARRGGGSGCYNCGELGHISKDCGIGGG
GGGGERRSRGGEGCYNCGDTGHFARDCTSAAGNDQRGATKGGNDGCTCGDVGHVARDCT
QKSVNGDQRGAVKGGNDGCTCGDVGHFARDCTQKVAAGNVRSGGGSGTCYSCGGVGH
IARDCATKRQPSRGYQCGSGHLARDCDQRSGGGGNDNACYKCGKEGHFAECSSVA*
>G770 (119..1069)

CCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGACACGCTGACAAGCTGACTCT
AGCAGATCTGGTACCGTCGACGGTTCCTTGGATTGGAGTAACTAAAGATCATATAAAAT
GGAACAAGGAGATCATCAGCAGCATAAGAAAGAAGAAGAAGCTTGGCCACCGGGTTTCAG
ATTTCAATCCGACGGATGAGGAGCTAATCTCATATTACTTGGTTAATAAGATTGCCGATCA
AACTTCACCGGAAAGCAATCGCTGACGTTGATCTTAACAAGTCCGAGCCATGGGAGCT
TCCTGAGAAGCGGAAAATGGGAGGAAAGAATGGTACTTTTTTAGCCTCCGGGACCGGAA
GTACCCGACCGGAGTGAGGACGAATAGGGCGACGAATACAGGATATTGGAACACAGG
AAAAGACAAAGAGATATTCAATAGCACAACTCGGAGTTGGTCCGGATGAAGAAGACTTT
GGTCTTTTACAGAGGACGAGCTCCTCGTGGGGAGAAGACTTGTGGGTCTGTCATGAGTA
TCGACTTCACTCCAAGTCTCATATAGAACCTCCAAGCAAGACGAGTGGGTAGTGTGTAG
AGTGTTCAGAAAACAGAAGCAACCAAGAAATACATAAGCACCAGTAGCAGCAGCACAAAG
TCATCACCACAACAACACACAAGAGCCTCAATACTATCAACCAACAACAATAATCCTAA
TACTCATCAGACCTCCTTCAACTCCCAACCGCATCTACAACACACCCGAGCCTCAATAT
TAACCAATCCCTCATGGCAAACGCCGTTACCTAGCTGAGCTCTCAAGAGTCTTCCGTGC
CTCTACAAGCACCACCATGGACTCTTCTCATCAGCAGCTAATGAACACACCCACATGCC
TGTCTCAGGGCTCAACCTCAACCTTGGCGGTGCACTGGTCCAGCCGCTCTCTGTTGTGTC
TCTTGAGGATGTTGCGCGGTAGTGCTTCTGTAATGGCGAAAACGGGTTTGAAATGT
GGAGATGAGCCAGTGCATGGACTTGGATGGATACTGGCCATCTTATTGATTGGTAATTGT
CAGTTTAAGTTATGTTTATATTGTTTCCATTTACTTGTGGTAAAACGATTTTGGTT
GTTCTTGCGAACCTCTAGACAGGCCTCGTACCGATCCTCTAGCTAGAGCTTTCGTTCTG
TATCATCGGTTTC

>G770 Amino Acid Sequence (domain in AA coordinates: 19-162)
MEQGDHQHKKKEEALPPGFRFHPTDEELISYYLVNKIADQNTGKAIADVDLNKSEPWE
LPEKAKMGGKEWYFFSLRDRKYPTGVRTNRTNTGYWKTGKDKEIFNSTSELVGMKKT
LVFYRGRAPRGEKTCWVMHEYRLHKSYSYRTSKQDEWVVCRVFKTEATKKYISTSSST
SHHHNNHTRASILSTNNNNPNYSSDLLQLPPLQPHPSLNINQSLMANAVHLAELSRVFR
ASTSTTMDSSHQQLMNYTHMPVSGNLNLGGLVQPPPVVSLDVDAAVSASYNGENGFN
VEMSQCMDLDGYWPSY*
>G858 (99..869)

CATAATCTCTCTCTATATCTCTTCTCTTCTTTTACCCTGTTTTTTTTTTCATTC
CACAGAGCCAGGTTGATTGATTTTGTATTATTCAGAGATATGGGAGAGGAAGGATTGAGA
TTAAGAAGATTGAGAATATCAACAGTCTGCAAGTCACTTTCTTAAGAGACGAAACGGTT
TGATCAAGAAGGCTAAAGAGCTTTCGATTCTCTGTGACGCCGAGGTTGCTCTTATCATCT
TCTCCAGCACCGGCAAGATTTACGATTTCTCCAGCGTCTGTATGGAGCAAATTCTTTCTA
GATATGGATACACTACTGCGTCCACTGAGCATAAACAACAAGAGAACACCAACTTCTAA
TTTGTGCTTCACATGGAAATGAAGCTGTGTTGCGAAATGATGATTCTGAAGGGGGAAC
TTGAAAGATTACAGCTTGCAATTGAGAGACTTAAGGGTAAGGAGCTTGAAGGTATGAGTT
TCCCGGATCTTATTTCTTCTTGAACCAAGTTGAACGAGAGCTTGATAGTGTCAAGGATC
AAAAGACACAAATCTGCTCAACCAGATTGAGAGATCCAGGATACAGGAGAAAAAGCAT
TGAAGAAAACCAATCTTGCGCAAACAGGTTGAGATGTTGGGAGAGGTTTCAGGACCA
AAGTGTGAATGAAAGGCCTCAAGATTCTAGCCCAGAAGCCGATCCCGAGAGCTCTTCAT

CAGAAGAGGATGAGAATGACAACGAGGAGCACCATTCCGACACTTCCTTGCACTTGGGGT
TGTCGTCGACGGGGTATTGCACAAAGAGAAAGCCGAAGATCGAACTGGTCTGCGATA
ACTCTGGGAGTCAAGTGGCTTCTGATTGATGGAATCGATTATTTTCTAATTCTGGTTGT
TTAGGGGTCTCTATGTGTCTTCTTGTCTTCTGGCTGTTCTTTTGCTTTATTTTCATCTCAAG
TAGAGTTTCTTAATGTTTTAGGTGGAACATTTTCCATAATCAAGAAGGGATTGATCAA
TCAATAACATTAGATTTTCTTAGTTAAAGACTTAAAGTTGCCACACACCACACCATATG
TGATTATGATGAATTTACATTTTATAAAAAAAAAAAAAAAAAAAAAAAAAA
>G858 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRGRIEIKKIENINSRQVTFKRRNGLIKKAKELSILCDAEVALIIFSSTGKIYDFSSV
CMEQILSRYGYYTASTEHHKQREHQLLICASHGNEAVLRNDDSMKGELERLQLAIERLKG
KELEGMSFPDLISLENQLNESLHSVVDQKTQILNLQIERSRIQEKKALEENQILRKQVEM
LGRSGPKVLNERPQDSSPEADPESSSSSEEDENDNEEHSDTSLQLGLSSTGYCTKRKKP
KIELVCDNSGSQVAD*
>G865 (282..920)
ATCCCCACTTGTGTTCATCACCAAGCCAAGCTCCATGTCCTAGTCACTCCACAGATTCC
CTATCATCATCAATTCGTTTCAAACCTTAGTTCCTTTCAAAGTCTTGATACATATATACACA
CACACCTATTATTCTCTTGGTGTGTTTGTGTGTACATATACGTGTGAGTACATACTTTG
TTGTAAGTGGATCGGAGGTATGGAAGGGACCGGTTCCACCGGAAACATCGGCGCGCG
CGGATGATAATTCTGCTCTTGGACGAGACTGATGTACCGCCATGGTCTCCGCTCTCAGCC
GTGTATAGAGAATCCGACAGACCCCGCGGTCAAACAAGAGCTTGATAAATCGGATCAAC
ATCAACCAGACCAAGATCAACCAAGAAGAAGACACTATAGAGGCGTAAGGCAGAGACCAT
GGGGTAAATGGCGCGCAGAAATCCGCGATCCAAAGAAAGCAGCCCGTGTCTGGCTCGGGA
CTTTTCGAGACGGCAGAGGAAGCTGCTTTAGCCTATGACCGAGCTGCCCTCAAATTCAAAG
GCACCAAGGCTAAACTGAACTTCCCTGAACGGGTCCAAGGCCCTACTACCACCACAACCA
TTTCTCATGCACCAAGAGGAGTTAGTGAATCCATGAACTCACCTCCTCCTCGACCTGGTC
CACCTTCACTACTACTTCTCGTGGCCAATGACTTATAACAGGACATACTTCAATACG
CTCAGTTGCTTACAGATAACAATGAGGTTGATTTATCATACTACACGTCGACTCTCTTCA
GTCAACCTTTTCAACGCCTTCTTCATCTTCTTCTTCTCCCAACAGACGCAGCAACAGC
AGCTACAACAACAACAACAGCAGCGTGAAGAAGAAGAGAAGAAATTATGGTTACAATTATT
ATAACTACCCAAGAGAATAATCTAATTATTATTGTTGGTTCGAATCAGTTTATAAATAGC
TATCATAGTTTTCATTTTGGTTTCCGTAACCTTTGTTGTCATGGAAAATATGAATGAACGA
GGGACATGTGTAACAATTTGTTTGTGTTTCGTAAATGTTAGTTGATTTGGATTGCTGA
AGTTTGATTTTCTGAGCATAAATCATTTGACGGTCAAAAAAAAAA
>G865 Amino Acid Sequence (domain in AA coordinates: 36-103)
MVSALSRVIENPTDPPVKQLDKSDQHQPDQDQPRRRHYRGVVRQRPWGKWAEEIRDPKKA
ARVWLGTFFETAEBAALAYDRAALKFKGTAKLNFPERVQGPTTTTTISHAPRGVSESMNS
PPRPGRPPSTTTTTPMTYNQDILQYAQLLTSNNEVDLSYYTSLFSQPFSTPSSSSSSS
QQTQQQQLQQQQQQREEEKNYGYNYNYNPRE*
>G872 (59..646)
CCGGAACAGAAATCCAATTAACCAAAACCGAATCGAACCGAACCGGAGTTTTTATCCAAT
GGTGAAGCAAGCGATGAAGGAAGAGGAGAAGAAGAGAAACACGGCGATGCAGTCAAAGTA
CAAAGGAGTGAGGAAGAGGAAATGGGGAAAATGGGTATCGGAGATCAGACTTCCACACAG
CAGAGAACGAATTTGGTTAGGCTCTTACGACACTCCCAGAAAGCGCGCGTGCTTTTCGA
CGCCGCTCAATTTTGTCTCCGCGGCGCGATGCTAATTTCAATTTCCCTAATAATCCACC
GTCGATCTCCGTAGAAAAGTCGTTGACGCCTCCGGAGATTGAGGAAGCTGCTGCTAGATT
CGCTAACACATTTCAAGACATTGTCAAGGGAGAAGAAGAAATCGGGTTTAGTACCCGGATC
CGAGATCCGACAGAGTCTCTTCTACATCTGCATCTGTTGCTACATCGACGGTGATTGA
TGATTTTTCGTTTTTGGATTGCTTCCGATGAATTTGCGGTTTGATTCTTCTCCGACGA
CTTCTCTGGCTTCTCCGGTGGTGATCGATTTACAGAGATTTTACCCATCGAAGATTACGG
AGGAGAGAGTTTATTAGATGAATCTTTGATTCTTTGGGATTTTGAATTCCCAAACATAA
TATTTTATTAGAGCGAAGTGTGAGATTTTCTTGGAGTCATGGAGAAATCTGGAGATTTT
TTGTAACACGGAGCTCCAATGACCCGGGAATTTCTTCTGTTTCGGATCCGAATTTGATGT
GGATCATATTACACCTATATTTTTCATTTTTTGTGTGTAAGAAAAATCGGATAAGAT
TCTAGTAATAAATGTAAAGTCCATTTTCAATTAATAAAAAAAAAAAAAAAAAA
>G872 Amino Acid Sequence (domain in AA coordinates: 18-85)
MVKQAMKEEEKRNTAMQSKYKGVKRKWKVSEIRLPHSRERIWLGSYDTPEKAARAF
DAAQFCLRGGDANFNFPNPPSISVEKSLTPPEIQEAAARFANTFQDIVKGEEESGLVPG

SEIRPESPSTASVATSTVDYDFSFLDLLPMNFGFDSFSDDFSGFSGGDRFTEILPIEDY
GGESLLDESLILWDF*

>G904 (1..1005)

atggaatctctcatcaatcccagccatggcggaggaaactacgattctcactcttcttctc
ctcgatagtctcaaaccaagcgtagtcatcattctcattctcctcatgactcttctc
atctccgtttccatttgcttctcctcctcgctgtctcaatcgctgtagccaccgctccgtt
ctcctcttttcatcttctccttccgtcgcaaccgtaacttccgattcccgcagattctct
ggacatcgagtctctccgaacagaacgggtcctcctgcttgattcgcttccgattttc
aaattctcctccgtcactcgccgatctagctccatgaattccggagattgcgccgtttgt
ttgtcgaaattcgaaccggaggatcagctccgtcttcttctctctgttgacgctttt
cacgccgattgtatcgatatctggctagtctctaaccagacttgctctctctgctgctct
cctctcttctgcttcagaatctgatctcatgaagtctctcgccgtcgctcggtcaaacaac
ggcggaggagaaaacagcttccgtctcgaaatcggatccatcagccgctcgctcgtaaaca
ccgattccagaatccggttgagcagcatcgaacttactcaatcggttgcgttcgattacata
gtagacgacgttagattcagaatctcagagtcaaatttcaaccgtggaaaacaggaagac
gcgactacaacaactgccacagcaacggcggttacgactaatccgagctcgttgaagct
agtttagcggcggtataggtaacgatggttctagaagctggctcaaggattacggtgac
agactctcacgaggtatctcgctcggtgcaatgtcggttagaagctctggttagattttt
actgggagtagtctcgaggtgaggaattgacggtgatggatttagaagcgaatcatgcc
ggagaagagataagtgagcttttccggtggctctcaggggtgtga

>G904 Amino Acid Sequence (domain in AA coordinates: 117-158)

MESLINPSHGGGNYDSSSSLDLSPSVLVIIILLMTLLISVSICFLRLNRCSHRSV
LPLSSSSSVATVTSDSRFRSGHRVSPETERSSVLDLPIFKFSSVTRRSSMNSGDCAVC
LSKFEPEDQLRLPLCCHAFHADCIDIWLVSNQTCPLCRSPLFASES DLMKSLAVVGSNN
GGGENSFRLEIGSISRRTPIPIESVEQHRYSIGSFDYIVDDVDSEISESNFNRGKQED
ATTTTATATAVTNTPTSFASLAADIGNDGRSRLKDYVDRLSRGISSRAMSFRSSGRFF
TGSSRRSEELTVMDLEANHAGEEISELFRWLSGV*

>G910 (1..1071)

ATGTTATGTATAATAAATGAGAATATGGAAAGAGTATGTGAGTTTGTAAAGCGTAT
AGAGCAGTGGTTTATTGTATAGCTGATACAGCAAATCTTTGTTTAAATGTGATGCAAAG
GTTCAATCAGCTAATTCACCTCTCGGGACGGCATTACGTACGGTTTATGTGATTCTGGT
AAGAATCAGCCTTGTGTTGTCGGATGTTTGGACCATAAAATGTTCTTTGCCATGGATGT
AATGATAAGTTTCATGGTGGTGGCTCTTCTGAGCATCGTAGAAGGGATTGAGGTGTTAT
ACGGGTTGTCTCTCTGCTAAAGATTTCGCGGTTATGTGGGGTTTTCGAGTTATGGATGAC
GATGATGATGTTTCGTTAGAGCAATCTTTTGAATGGTAAACCTAAGGTGCAAAGAGAA
GGTGGTTTTATCTTGAACAGATTCTTGAATTGGAGAAGGTTTCAGCTCAGGGAAGAGAAT
GGTAGTTCTTCTTGACAGAACGAGGTGATCCATCTCCATTGGAGCTTCTTAAGAAACCC
GAAGAACAGTTAATCGATCTTCCGCAGACCGGAAAGAGCTGGTTGTTGATTTTTCACAC
TTGTCTCTCATCTTCCACTTGGTGATTCTTTTGGGAATGCAAAGTCCATACAATAAG
AACAATCAGTTGTGGCATCAAAATATACAAGACATTGGAGTATGTGAAGATACAATCTGC
AGTGACGATGACTTCCAAATACCTGACATTGATCTCACTTTCCGGAACCTTGAAGAGCAA
TTTGGAGCTGATCTTGAGCCAATTGCAGATAGTAACAACGTGTTCTTTGTTCTTCCCTT
GACAAATCACATGAGATGAAGACATTTCTTCTTCAATCAATAATCCCATATTTGCACCT
AAACCAGCTTCATCACTATCTCATTTCTCAAGCAGTGAACCGGATAACCCCTTATAGTCAC
TCAGAGGAAGTAATCTCATTTTGTCCCTCCCTCTCTAACAATACACGTCAAAGGTCATC
ACAAGGCTCAAGGAGAAGAAGAGCAAGAGTGGAGGAGAAAAAGCTTAA

>G910 Amino Acid Sequence (domain in AA coordinates: 14-37, 77-103)

MLCIIIIENMERVCFCKAYRAVVYCIADTANLCLTCDKVVHSANSLSGRHLRTVLCDSG
KNQPCVVRFCDFHKMFLCHGKNDKFHGGSSSEHRRDLRCYTGCPKADFVMMWGFVMD
DDVSLLEQSFMRVVKPVQREGGFLEQILELEKVLREENGSSSLTERGDPSPLELPKPP
EEQLIDLPTQKELVVDFSHLSSSSLTGDSFWECKSPYNKNNQLWHQNIQDIGVEDTIC
SDDDFQIPDIDLTFRNFEEQFADPEPIADSNVFFVSSLDKSHMKTFSSSFNNPIFAP
KPASSTISFSSSETDNPYSHSEEVISFCPSLSNNTQKVITRLKEKKRARVEEKKA*

>G912 (20..694)

CATCTTATCCAAAGAAAAAATGAATCCATTTTACTCTACATTCCCAGACTCGTTTCTCTC
AATCTCCGATCATAGATCTCCGGTTTCAGACAGTAGTGAGTGTACCAAAGTTAGCTTC
AAGTTGTCCAAAGAACGAGCTGGGAGGAAGAAGTTTCGTGAGACACGTATCCGATTTA

CAGAGGAGTTCGTCAGAGGAATTCTGGTAAATGGGTTTGTGAAGTTAGAGAGCCTAATAA
GAAATCTAGGATTTGGTTAGGTACTTTCCGACGGTTGAAATGGCTGCTCGTGCTCATGA
TGTTGCTGCTTTAGCTCTTCGTTGGTGGCTCTGCTTGTCTCAATTTGCTGATTCTGCTTG
GCGGCTTCGTATTCCTGAGACTACTTGTCTTAAGGAGATTAGAAAAGCTGCGTCTGAAGC
TGCAATGGCGTTTCAGAATGAGACTACGACGGAGGGATCTAAACTGCGGCGGAGGCAGA
GGAGGCGGAGGGGAGGGGGTGAAGGAGGGGAGAGGAGGGCGGAGGAGCAGAATGGTGG
TGTGTTTTATATGGATGATGAGGCGCTTTTGGGGATGCCCAACTTTTTTGAGAATATGGC
GGAGGGGATGCTTTTGCCGCCGCCGGAAGTTGGCTGGAATCATAACGACTTTGACGGAGT
GGGTGACGTGTCACTCTGGAGTTTGGACGAGTAATTTTTTGGCTCTTTTTCTGGATAATA
AGTT

>G912 Amino Acid Sequence (domain in AA coordinates: 51-118)
MNPFFYSTFPDSFLSISDHRSPVSDSSECSPLASSCPKKRAGRKKFRETRHPIYRGVRQR
NSGKWVCEVREPNNKSRWLGTFTVEMAARAHDAALALRGRSACLNFAWSAWRLRIPE
TTCPEIQKAASEAAMAFQNETTEGSKTAAEAEEAAGEGVREGERRAEEQNGGVFYMD
EALLGMNFFENMAEGMLLPPEVGNHNDFFDGVGDSLSWSFDE*

>G920 (114..1154)
AAAAATCTATTTCTTCTCTTCCACTATATTACAACATTTCTTCATTCTCAAATCATC
ATACTAAAAACCTAAAAAAGTTACATATTCAATTGTATCTTTGTGAGAAAAAATGGATT
CGAATAGTAACAACACGAAATCCATAAAGAGAAAAAGTTGTGACCAACTTGTGGAAGGCT
ATGAATTCGCTACTCAGCTTCAGCTTCTCCTTTCTCATCAACACTCTAACCAGTACCACA
TCGATGAGACCCGCTTGTGTTCCGGGTCGGGTTCAAGTTCCGGTGGTCCAGATCCCGTTG
ATGAGCTCATGTCTAAGATCTTGGGATCTTTCCATAAAACTATATCGGTTCTTGATTCTT
TTGATCCCGTCGCGCTCTCTGTCCCATCGCCGTCGAGGGTTTCATGGAATGCTTCATGTG
GGGATGATTGCGCGACTCCGGTGAGTTGCAACGGTGGAGATTCCGGTGAGAGTAAGAAGA
AGAGATTAGGGGTTGTTAAGGTAAGGTAAGAGGATGCTACACTAGAAAAGACGAGATCACATA
CAAGGATCGTGGAAGCTAAAGTTCTGAAGACAGATATGCTTGAGGAAATATGGACAAA
AGGAGATTCTTAATACCACATTCCCAAGAAGTTACTTTAGATGCACACACAAGCCAAACGC
AAGGATGCAAAGCAACAAGCAAGTTCAGAAACAGGATCAAGATTCTGAGATGTTCCAAA
TCACATACATTGGCTACCACACATGCACTGCCAATGACCAACGCACGCGAAGACCGAGC
CTTTTGATCAAGAAATCATTATGGATTCCGAAAAGACATTGGCTGCTAGCACTGCTCAGA
ACCATGTCAATGCTATGGTGCAAGAGCAAGAGAACAACACCAGCAGTGTGACAGCAATAG
ACGCAGGCTAGGGTTGTTAAGGAGGAACAAAATAACAATGGTGATCAGAGTAAGATTATTATG
AGGGCTCTTCGACAGGTGAGGACTTGTCAATTGGTTTGGCAAGAGACGATGATGTTTGATG
ATCATCAAAATCACTACTATTGTGGTGAAACCAGTACTACTTCTCATCAATTTGGTTTCA
TCGACAACGATGATCAGTTTTCCTCCTTCTTCGACTCATATTGTGCTGATTATGAAAGAA
CAAGTGCTATGTGAACATCCAAATCTGGAATGATGAATCAGCACTAGGTCTTCTCTTTGA
GTATGTCTAGTTTAATGTAATATTTTGTGTATGTTTGATAAACAACCATATATACTT
CTCTTTTACACCAAAAAAAAAAAAAAAAAAAAAA

>G920 Amino Acid Sequence (domain in AA coordinates: 152-211)
MDSNSNNTKSIKRKVVDQLVEGYEFATQLQLLLSHQHSNQYHIDETRLVSGSGSVSGGPD
PVDELMSKILGSFHKTIISVLDSFDPVAVSVPIAVEGSWNASCDDSATPVSCNNGDSGES
KKKRLGVGKGRGCYTRKTRSHTRIVEAKSSEDRYAWRKYQKEILNTTFPRSYFRCTHK
PTQGCATKQVQKQDQDSEMFQITYIGYHTCTANDQTHAKTEPFQDEIIMDSEKTLAAT
AQNHVNAMVQEQENNTSSVTAIDAGMVKEEQNNNGDQSKDYEGSSSTGEDLSLVWQETMM
FDDHQNHYYCGETSTTSHQFGFIDNDDQFSSFFDSYCADYERTSAM*

>G939 (9..1565)
CAGATTCTATGGATATGTATAACAACAATATAGGGATGTTCCGGAGTTTAGTTTGTAGCT
CGGCGCCTCCATTTACAGAGGGACATATGTGTTCTGATTGCGATACGGCTTTGTGCGATG
ATCTGAGTAGTGATGAGGAAATGGAATAGAGGAGCTTGAGAAGAAGATCTGGAGAGACA
AGCAGCGTTTAAAGCGGCTCAAGGAAATGCGGAAGAACGGTCTAGGAACAAGATTGTTGT
TGAAGCAGCAACATGATGATTTTCCAGAGCACTCTAGTAAGAGAACCATGTACAAGGCAC
AAGATGGGATCTTGAAGTACATGTCGAAGACAATGGAGCGATATAAAGCTCAAGGTTTGT
TTTATGGGATTGTGTTAGAGAATGGGAAAACGGTAGCGGGATCTTCTGATAATCTCCGTG
AATGGTGGAAGACAAAGTGAGGTTTGTATAGGAACGGCCAGCTGCTATAATCAAGCACC
AAAGGGATATCAATCTTCTGATGGAAGTATTGAGGTTCTGAGGTTGGGGATTCTACCG
CACAGAAGTTGCTTGAGCTTCAAGATACTACTCTTGGAGCTCTGTTATCGGCTCTGTTTC
CTCACTGCAACCTCTCAGAGGCGGTTTCCGTTGGAGAAAGGCGTGACACCGCCATGGT

GGCCAACGGGGAAAGAAGATTGGTGGGATCAACTGTCTTTACCCGTTGATTTTCGAGGTG
TTCCGCCACCTTACAAGAAGCCTCATGATCTCAAGAAGCTGTGGAAAATTGGTGTGTTGA
TTGGTGTAATCAGACATATGGCTTCTGACATTAGCAACATACCCAATCTCGTGAGACGGT
CTAGAAGTTTGCAGGAGAAAATGACGTCAAGAGAAGGCGCTTTATGGCTCGCTGCTCTTT
ACCGAGAAAAGGCTATTGTTGATCAAATAGCCATGTCTAGAGAAAACAACAACACTTCTA
ACTTTCTTGTTCCTGCAACCGGTGGAGACCCAGATGTTTTGTTTCTGAATCTACAGACT
ATGATGTTGAATGATGTTGGTGGCACTCATCGGACCAATCAGCAGTATCCTGAATTTGAAA
ACAACACTAACTGTGTTTACAAGAGAAAAGTTTGAAGAAGATTTTGGGATGCCAATGCATC
CAACACTCCTAAATGTGAGAACAGTCTCTGTCTTATAGCCAACCACATATGGGATTTCT
TTGACAGGAACCTAAGAGAGAATCACCAATGACTTGTCTTATAAAGTCACTTCTCTCT
ACCAACCAACTAAACCCATGGTATGACGGGTTTAAATGGTTCCTTGTCCGGATTATAACG
GGATGCAGCAGCAGGTTTCAAGACCACTTCAAGACCACTTCAATCATCCCAACGATCTCTACA
GACCAAAAGCTCCACAAAGAGGCAACGATGACTTGGTTGAGGATTTGAATCCTTCTCTCT
CGACGCTGAATCAGAACTTGGTTTAGTCTTACCTACTGACTTCAATGGAGGTGAGGAAA
CAGTAGGAACAGAGAACAATCTGCATAATCAAGGGCAAGAGTTGCCACATCTTGGATTC
AGTAAAGAAAGCTTCAAGAGTTTCTTTTATGTTTCTAGTCTTTATAGCTTGTCTCTT
GCTTATCTCTCATTAAACACAGTTTTGTATCTCTCCATTTATAGCCCATGTAGCAATG
GAGAAGATTAGGTTTATAATAAGTTAATAACCAAATTCAA
>G939 Amino Acid Sequence (domain in AA coordinates: 97-106)
MDMYNNIGMFRSLVCCSAPPFTEGHMCSDSHTALCDDLSSDEEMIEIELEKKIWRDKQR
LKRLEKEMAKNGLGTRLLKQHHDFPEHSSKRTMYKAQDGILKYSKTMERYKAQGFVYG
IVLENGKTVAGSSDNLREWWDKVRFDNRGPAAILKHQDINLSDGSDSGSEVGDSTAQK
LLELQDITLGLALSALFPHCNPPQRRFPLEKGVTPPWPTGKEDWWDQLSLPVDVFRGVPP
PYKKPHDLKKLWKIGVLIGVIRHMASDISNIPNLVRRSRLQEKMTSREGALWLAALYRE
KAIVDQIAMSRNNNTSNFLVPATGGDPDVLFPSTDYDVELIGGTHRTNQOYPEFENNY
NCVYKRKFEEFDMFMPHPTLLTCENSLCPYSQPHMGFLDRNLRENHQMTCPIKVTSTFYQP
TKPYGMTGLMVPCPDYNGMQQVQSFQDQFNHPNDLYRPAKQGRNDDLVEDLNPSPTL
NQLGLVLPDFTDNGGEETVGTENNLHNQQLPTSWIQ*
>G963 (1..897)
ATGAGTTTGCCTCCAGGATTCAGGTTTCATCCCACTGATGAAGAACTGGTGGCTTACTAT
CTTGATAGGAAGGTCAACGCCAAGCCATTGAGCTCGAGATCATCCAGAAGTTGATCTT
TATAAATGCGAGCCATGGGACTTGCCTGAAAAGTCACTTTTGCCTGGGAAACGACATGGAA
TGGTACTTTTACAGCACAGGGATAAGAAGTATCCAAATGGCTCTAGGACGAACCGTGCG
ACCGGAGCGGTTTACTGGAAGGCCACGGGGAAAGATCGTACAGTAGAATCAAAGAAGATG
AAGATGGGAATGAAGAAGACACTGGTTTATTATAGAGGAAGGGCTCCTCATGGCCTTCGT
ACTAATTGGGTGATGCATGAATATCGTCTCACGCACGCTCCTTCTCTCTTGAAGGAG
TCGTATGCATTTGTGCCGAGTGTGTTAAGAAGAACATACAAATCCAAAGAGAAAAGGGGAA
GAAGAAGAAGCAGAAGAGAGAGCACTAGTGTAGGAAAAGAAGAGGAAGAAGAAAAGGAG
AAGAAGTGGAGAAAATGTGATGGTAATTATATTGAAGACGAGAGCTTGAAAAGAGCATCC
GCGGAGACATCTTCATCAGAGCTAACTCAAGGGGTCCTTTTAGACGAAGCAAAACAGCTCA
TCCATATTTGCTCTTCTATTTCTCATCTTCTCTTCTGACGATCATGATCATCTTTCTCA
AACTATTCTCATCAGCTTCCATATCATCTCTCTTCACTCCAAGATTTCCCTCAACTT
TCTATGAACGAAGCAGAGATTATGTCAATCCAACAAGACTTTCAATGCAGAGACTCTATG
AACGGGACACTTGACGAAATCTTCTCTTCTCCGCCACTTTCCCGCTTCCCTTTGA
>G963 Amino Acid Sequence (domain in AA coordinates: TBD)
MSLPPGFRFHPTDEELVAYYLDKRVNGQAELEIIEPVDLYKCEPWDLPEKSFLPGNDME
WYFYSTRDKKYPNGSRTNTRATAGYWKATGKDRTVESKKMKMGKTLVYYRGRAPHGLR
TNWVMHEYRLTHAPSSSLKESYALCRVFKKNIQIPKRKGEEEEEESTSVGKEEEEEKE
KKWRKCDGNYIEDESLKRASAEETSSSELQGVLLDEANSSSIFALHFSSSLDDHDHLFS
NYSHQLPYHPPLQLQDFPQLSMNEAEIMSIQQDFQCRDSMNGTLDEIFSSSATFPASL*
>G979 (60..1352)
CCTCTGAGGAATCAAATCACTCACACTCCAAAAAATCTAACTTTCTCAGAGTTTAA
TGAGAAGCGCTTAACCACTTCCACTTGTCTTCTTCTCCATCTTCTCTGTTTCTTCTT
CTACTACTACTTCTCTCTTCTCAGTGGAGGCTCCAAGGCCTAAACGAGCCAAAAGGG
CTAAGAAATCTTCTCTTCTGGTGATAAATCTCATAACCCGACAAGCCCTGCTTCTACCC
GACGCAGCTCTATCTACAGAGGAGTCACTAGACATAGATGGACTGGGAGATTGAGGCTC
ATCTTTGGGACAAAAGCTCTTGAATTTCGATTAGAACAAGAAAGGCAACAAGTTTATC

TGGGAGCATATGACAGTGAAGAAGCAGCAGCACATACGTACGATCTGGCTGCTCTCAAGT
ACTGGGGACCCGACACCATCTTGAATTTTCCGGCAGAGACGTACACAAAGGAATTGGAAG
AAATGCAGAGAGTGACAAAGGAAGAATATTTGGCTTCTCTCCGCCAGAGCAGTGGTT
TCTCCAGAGGCGTCTCTAAATATCGCGGCGTCGCTAGGCATCACCACAACGGAAGATGGG
AGGCTCGGATCGGAAGAGTGTGTTGGGAACAAGTACTTGTACCTCGGCACCTATAATACGC
AGGAGGAAGCTGCTGCAGCATATGACATGGCTGCGATTGAGTATCGAGGCGCAAACGCGG
TTACTAATTTTCGACATTAGTAATTACATTGACCGGTTAAAGAAGAAAGGTGTTTTCCCGT
TCCCTGTGAACCAAGCTAACCATCAAGAGGGTATTCTTGTGTAAGCCAAACAAGAAGTTG
AAACGAGAGAAGCGAAGGAAGAGCCTAGAGAAGAAGTGAAACAACAGTACGTGGAAGAAC
CACCGCAAGAAGAAGAAGAGAAGGAAGAAGAGAAGCAGAGCAACAAGAAGCAGAGATTG
TAGGATATTGAGAAGAAGCAGCAGTGGTCAATTGCTGCATAGACTCTTCAACCATAATGG
AAATGGATCGTTGTGGGGACAACAATGAGCTGGCTTGGAACTTCTGTATGATGGATACAG
GGTTTTCTCCGTTTTTGTACTGATCAGAATCTCGCGAATGAGAATCCCATAGAGTATCCGG
AGCTATTCAATGAGTTAGCATTTGAGGACAACATCGACTTCATGTTTCGATGATGGGAAGC
ACGAGTGTGTTGAACCTTGGAAATCTGGATGTTGCGTGGTGGGAAGAGAGAGCCCACCCCT
CTTCTTCTTACCATTGTCTTGCTTATCTACTGACTCTGCTTCATCAACAACAACAACAA
CAACCTCGGTTTTCTGTAACTATTTGGTCTGAGAGAGAGAGCTTTGCCTTCTAGTTTGAA
TTTCTATTTCTCCGCTTCTTCTTTTCTTTTCTTTTGTGGGTTCTGCTTAGGGTTTG
TATTTACGTTTCCGGCTTGTTCGTTGGTTCTGAATAATCAATGTCTTTGCCCTTTTNN
AANGNTNCAAGNTNAAANAAAAAAAAAAAA

>G979 Amino Acid Sequence (domain in AA coordinates: 63-139,165-233)

MKRLTSTSCSSPSSSVSSSTTTSSPIQSEAPRPKRAKRAKSSPSGDKSHNPTSPAST
RRSSIYRGVTRHRWTGRFEHLWDKSSWSN IQNKGKQVYLGA YDSEEAHAHTYDLAALK
YWGPDITLNFPAETTYTKELEEMQVRVTKEEYLA SLRRQSSGFSRGVSKYRGVARHHHNGRW
EARIGRVFGNKYLYLGTINTQEBAAAAYDMAAIEYRGANAVTNFDISNYIDRLKKKGVP
FPVNQANHQEGILVEAKQEVETREAKEEPREEVKQYVEEPPQEEEEKEEEKAEQQAEEI
VGYSEEA VVNCCIDSSTIMEMDRCDGNNELAWNFCMMDTGFSPFLTDQNLANENPIEYP
ELFNELAFEDNIDFMFDDGKHECLNLENLDCCVVGRESPPSSSSPLSCLSTDSASSTTTT
TTSVSCNYLV*

>G987 (1..4011)

ATGGGTTCTTACTCAGCTGGCTTCCCTGGATCCTTGGACTGGTTTGATTTTCCCGGTTTA
GGAAACGGATCCTATCTAAATGATCAACCTTGTAGATATTGGATCTGTTCTCTCCTCCT
CTAGACCCATATCCTCAACAGAACTTGGCTTCTGCGGATGCTGATTTCTCTGATTCTGTT
TTGAAGTACATAAGCCAAGTTCTTATGGAAGAGGACATGGAAGATAAGCCTTGTATGTTT
CATGATGCTTTTATCTCTTCAAGCAGCTGAGAAGTCTCTCTATGAAGCTCTCGGCGAGAAG
TACCCGGTTGATGATTCTGATCAGCCTCTGACTACTACTACTAGCCTTGCTCAATTGGTT
AGTAGTCCTGGTGGTTCTTCTTATGCTTCAAGCACCACAACCACTTCTCTGATTCAAA
TGGAGTTTTGATTGTTTGGAGAATAATAGGCCTTCTTCTTGGTTGCAGACACCGATCCCG
AGTAACTTCATTTTTCAGTCTACATCTACTAGAGCCAGTAGCGGTAACGCGGTTTTCGGG
TCAAGTTTTAGCGGTGATTTGGTTTCTAATATGTTAATGATACTGACTTGGCGTTACAA
TTCAAGAAAGGGATGGAGGAAGCTAGTAAATTCCTTCTAAGAGCTCTCAGTTGGTTATA
GATAACTCTGTTCTTAACAGATTAAACCGAAAGAAGAGCCATTGGCGCGAAGAAGAACAT
TTGACTGAAGAAAGAAGTAAGAAACAATCTGCTATTTATGTTGATGAAACTGATGAGCTT
ACTGATATGTTTGACAATATTCTGATATTTGGCGAGGCTAAGGAACAACCTGTATGCATT
CTTAACGAGAGTTTCCCTAAGGAACCTGCGAAAGCTTCAACGTTTAGTAAGAGTCCTAAA
GGCGAAAAACCGGAAGCTAGTGGTAACAGTTATACAAAAGAGACACCTGATTTGAGGACA
ATGCTGGTTTTCTTGTGCTCAAGCTGTTTCGATTAAACGATCGTAGAACTGCTGACGAGCTG
TTAAGTCGGATAAGSCAACATTCTTCATCTTACGGCGATGGAACAGAGAGATTGGCTCAT
TATTTTGCTAACAGTCTTGAAGCACGTTTGGCTGGGATAGGTACACAGTTTTTACTGCC
TTGTCTTCCAAGAAACATCTACTTCTGACATGTTGAAAGCTTATCAGACATATATATCA
GTCTGTCCGTTTCAAGAAAATCGCAATCATATTCGCCAACCATAGTATTATGCGGTTGGCT
TCAAGTGCTAATGCCAAAACCATCCACATCATAGATTTTGGAAATATCTGATGGTTTCCAG
TGGCCTTCTCTGATTCTGACTTGCTTGGAGACGTGGTTTCTCTGTAAGCTTCCGATA
ACCGGTATAGAGTTGCTTCAACGCTGGTTTGTAGACAGCGGAGGAGTTATTGAGACTGGT
CGTCGCTTGGCTAAGTATTGTGAGAAGTTCAATATTCCGTTTGGAGTACAATGCGATTGCG
CAGAAATGGGAATCAATCAAGTTGGAGGACTTGAAGCTAAAAGAAGGCGAGTTTGTGCG
GTAAACTCTTTATTTCCGGTTTAGGAATCTTCTAGATGAGACGGTGGCAGTGCATAGCCCCG

AGAGATACGGTTTTGAAGCTGATAAGGAAGATAAAGCCAGACGTGTTTCATCCCCGGGATC
CTCAGCGGATCCTACAAACGCGCCTTTCTTTGTGACGAGGTTTAGAGAAGTTCTGTTTCAT
TACTCATCTCTGTTTGACATGTGTGACACGAATCTAACACGGGAAGATCCAATGAGGGTT
ATGTTTGAGAAAGAGTTCTATGGGCGGGAGATCATGAACGTGGTGGCGTGTGAGGGGACG
GAGAGAGTGGAGAGGCCAGAGAGTTATAAGCAGTGGCAGGCGAGGGCGATGAGAGCCGGG
TTTAGACAGATTCGGCTGGAGAAGGAAGTCTAGTTTCAGAACTGAAGTTGATGGTGGAAAGT
GGATACAAACCCAAAGAGTTTGATGTTGATCAAGATTGTCACTGGTTGCTTCAGGGCTGG
AAAGGTAGAATTGTATACGGTTCATCTATTGGGTTCCCTTTCTTTTCTATGTGGGCAGA
GCAACTAGGGTTTGATCATGGATCCAACTTCTCTGAATCTCTAAACGGCTTTGAGTAT
TTTGATGGTAACCTAATTTGCTTACTGATCCAATGGAAGATCAGTATCCACCACCATCT
GATACTCTGTTGAAATACGTGAGTGAGATTCTTATGGAAGAGAGTAATGGAGATTATAAG
CAATCTATGTTCTATGATTCAATTGGCTTTACGAAAACTGAAGAAATGTTGCAGCAAGTC
ATTACTGATTCTCAAAATCAGTCCTTTAGTCTCTGCTGATTCAATTGATTACTAATCTTGG
GATGCAAGCGGAAGCATCGATGAATCGGCTTATTCCGCTGATCCGCAACCTGTGAATGAA
ATTATGGTTAAGAGTATGTTTAGTGATGCAGAATCAGCTTTACAGTTTAAGAAAGGGGTT
GAAGAAGCTAGTAAATTCCCTTCCCAATAGTGATCAATGGGTTATCAATCTGGATATCGAG
AGATCCGAAAGGCGGATTCGGTTAAAGAAGAGATGGGATTGGATCAGTTGAGAGTTAAG
AAGATCATGAAAGGGATTGAGGAAGTTAGGAGTAGTAAGCAATTTGCTAGTAATGTA
GAAGATAGTAAGGTTACAGATATGTTTGATAAGGTTTGGCTTCTTGACGGTGAATGCGAT
CCGCAACATTGTTAGACAGCGAGATTCAAGCGATTCCGAGTAGTAAGAACATAGGAGAG
AAAGGGAAGAAGAAGAAGAAGAAGAGTCAAGTGGTGTATTTTCGTACACTTCTCACT
CATGTGTCACAAGCCATTTCACAGGAGATAAAACCACGGCTCTTGAGTTTCTGTTACAG
ATAAGGCAACAGTCTTCGCTCTCGGTGACGCGGGGCAAGACTAGCTCATTTGTTTCGCT
AACCGCTTGAAGCTCGTCTACAGGGAAGTACCGGCTCTATGATCCAGACTTATTACAAT
GCTTTAACCTCGTCTGTTGAAGGATACTGCTGCGGATACAATTAGAGCGTATCGAGTTTAT
CTTTCTTCGCTCCGTTTGTACCTTGATGATTTCTTCTCCATCTGGATGATTCTTGAT
GTGGCTAAAGATGCTCCTGTTCTTCATATAGTTGATTTGGGATTCTATACGGGTTTCAA
TGGCCGATGTTTATTAGTCTATATCAGATCGAAAAGATGTACCGCGGAAGCTGCGGATT
ACTGGTATCGAGCTTCCCTCAGTGCAGGTTTTCGGCCCGCGGAGCGAATAGAGGAGACAGGA
CGGAGATTGGCTGAGTATTGTAAACGGTTTAAATGTTCCGTTTGAGTACAAAGCCATTGCG
TCTCAGAACTGGGAAACAATCCGGATAGAAGATCTCGATATACGACCAAACGAAGTCTTA
GCGGTTAATGCTGGACTTAGACTCAAGAACCCTCAAGATGAAACAGGAAGCGAAGAGAAT
TGCCCCGAGAGATGCTGTCTTGAAGCTAATAAGAAACATGAACCCGGAGCTTTTCATCCAC
GCGATTGTTCAACGGTTCATTCAACGCACCTTCTTTATCTCGCGGTTTAAAGAAGCGGTT
TACCATTACTCCGCTCTCTTCGACATGTTTGATTCGACGTTGCCTCGGGATAACAAAGAG
AGGATTAGGTTTCGAGAGGGAGTTTACGGGAGAGAGGCTATGAACGTGATAGCGTGCGAG
GAAGCTGATCGAGTGGAGAGGCTGAGACTTACAGGCAATGGCAGGTTAGAATGGTTAGA
GCCGGGTTTAAAGCAGAAAACGATTAAGCCTGAGCTGGTAGAGTTGTTTAGAGGAAAGCTG
AAGAAATGGCGTTACCATAAAGACTTTGTGGTTGATGAAAATAGTAAATGGTTGTTACAA
GGCTGGAAAGGTCGAACCTCTCTATGCTTCTTCTGTTGGGTTCCCTGCCTAG
>G987 Amino Acid Sequence (domain in AA coordinates: 428-432,704-708)
MGSYSAGFPGLDWFDFPGLGNGSYLNDQPLLDIGSVPPPLDPYPQONLASADADFSDSV
LKYISQVLMEEDMEDKPCMFHDALSLQAAEKSLYEALGEKYPVDDSDOPLTTTTSLAQLV
SSPGGSSYASSTTTSSDSQWSFDCLNNRPSSWLQTPIPSNFIFQSTSTRASSGNAVFG
SSFGDLVSNMFNDTDLALQFKGMEEASKFLPKSSQLVIDNSVNPRLTGKKSHWREEEH
LTEERSKKQSAIYVDETDELTFMFDNILIFGEAKEQPVCIILNESFPKEPAKASTFSKSPK
GEKPEASGNSYTKETPDRLTMLVSCAQAVSINDRRTADELLSRIRQHSSSYGDTERLAH
YFANSLEARLAGIGTQVYTALSSKKTSTSDMLKAYQTYISVCPFKKIAIIFANHSIMRLA
SSANAKTIHIIDFGISDGFQWPSLIHRLAWRRGSSCKLRITGIELPQRGFRPAEGVIETG
RRLAKYCQKFNIPFEYNALAQWESIKLEDLKLKEGEFVAVNSLFRFRNLDETVAHVSP
RDTVLKLRIRKIPDVFIPIGLSGSYNAPFFVTRFREVLFHYSSLFDMCDTNLTREDPMRV
MFEKBFYGREIMNVVACEGTERVERPESYKQWQARAMRAGFRQIPLEKELVQKLKLMVES
GYKPKFEDVDQDCHWLLQGWKGRIVYGSSIWVPPFFYVGRATRVLIMDPNFSLSNGFEY
FDGNPNLLTDPMEDQYPPPSDTLLKYVSEILMEESNGDYKQSMFYDSLALRKTEMLQQV
ITDSQNQSFSPADSLITNSWDASGSIDESAYSADPQVNEIMVKSMSFSDAESALQFKGV
EEASKFLPNSDQWVINLDIERSERRDSVKEEMGLDQLRVKKNHERDFEVRSSKQFASNV
EDSKVTDMFDKVLILLDGECDPQTLLDSEIQAIRSSKNIGEGKKKKKKKSQVVDFTLLT

HCAQAISTGDKTTALEFLQIRQSSPLGDAGQRLAHCFANALEARLQGSTGPMIQTYYN
ALTSSLKDTAADTIRAYRVYLSSSPFVTLMYFFSIWMILDVAKDAPVLHIVDFGILYGFQ
WPMFIQSISDRKDVPRKLRTIGIELPQCGFRPAERIEETGRRLAEYCKRFNVPFEYKAIA
SQNWETIRIEDLDIRPNEVLAVNAGLRRLKLNQDETGSSENCPRDAVLKLRNMNPDVFIH
AIVNGSFNAPFFISRFKEAVYHYSALFDMFDSTLPRDNKERIRFEREFYGREAMNVIACE
EADRVERPETYRQWQVRMVRAGFKQKTIKPELVELFRGKLKKWRYHKDFVVDENSKWLLQ
GWKGRITLYASSCWVPA*

>G993 (6..1091)

CAAATATGGAATACAGCTGTGTAGACGACAGTAGTACAACGTCAGAATCTCTCTCCATCT
CTACTACTCCAAAGCCGACACGACGACGAGAGAAAGAACTCTCTTCTCCGCCGCGACGT
CGATGCGTCTCTACAGAATGGGAAGCGGCGGAAGCAGCGTCGTTTGGATTACAGAGAACG
GCGTCGAGACCGAGTCACGTAAGCTTCCCTCGTCGAAATATAAAGCGTTGTGCCTCAGC
CTAACGGAAGATGGGGAGCTCAGATTTACGAGAAGCATCAGCGAGTTTGGCTCGGTACTT
TCAACGAGGAAGAAGAAGCTGCGTCTCTTACGACATCGCCGTGAGGAGATTCCGCGGCC
GCGACGCGCTCACTAATCTCAAATCTCAAGTTGATGGAACGACGCCGAATCGGCTTTTC
TTGACGCTCATTCTAAAGCTGAGATCGTGGATATGTTGAGGAAACACACTTACGCCGATG
AGTTTGAAGCAGAGTAGACGGAAGTTTGTAAACGGCGACGGAACCGCTCTGGGTGGAGA
CGGCGACGTACGGAACGACGCTGTTTGTAGAGCGCGTGAGGTTTGTTCGAGAAGACTG
TTACGCCGAGCGACGTGCGGAAGCTGAACCGTTTAGTGATACCGAAACAACACGCGGAGA
AGCATTTTCCGTTACCGCGATGACGACGCGGATGCGGGATGAATCCGTCTCCGACGAAAG
GCGTTTTGATTAACTTGGGAAGATAGAACAGGGAAGTGTGGCGGTTCCGTTACAGTTACT
GGAACAGCAGTCAAAGTTACGTGTTGACCAAGGGCTGGAGCCGTTTCGTTAAAGAGAAGA
ATCTTCGAGCCGGTGATGTGGTTTGTTCGAGAGATCAACCGGACGACCGGCAATTGT
ATATCCACTGGAAGTCCGGTCTAGTCCGGTTCAGACTGTGGTTAGGCTATTCCGAGTCA
ACATTTTCAATGTGAGTAACGAGAAACCAAACGACGTGCGAGTAGAGTGTGTTGGCAAGA
AGAGATCTCGGAAGATGATTTGTTTTCGTTAGGGTGTTCGAAGAAGCAGCGGATTATCA
ACATCTTGAGCAAATCTTTTTTTTGGTTTTTTTCTCAATTTGTTTCTCTTTTCA
ATATTTTGTATTGAAATGACAAGTTGTAAATTAGGACAAGACAAGAAAAAATGACAACTA
GACAAAATAGTTTTTTGTTAAAAA

>G993 Amino Acid Sequence (domain in AA coordinates: 69-134)
MEYSCVDDSSSTSESLSISTTPKPTTTEKKLSSPPATSMRLYRMGSGGSSVVLDSNGV
ETESRKLPSKSKYKGVVPQNGRWGAQIYEKHQVRVWLGTFNEEEAASSYDIAVRRFRGRD
AVTNFKSQVDGNDAESAFDLAHSKAEIVDMLRKHTYADEFEQSRKRVNGDGKRSGLETA
TYGNDAVLRAREVLFKFTVPSDVGLNRLVIPKQHAEKHFPLPAMTTAMGMNPSPTKGV
LINLEDRTGKVVFRFRYSYWNSSQSYVLTKGWSRFVKEKNLRAGDVVCFERSTGPDRQLYI
HWKVRSSPVQTVVRLFGVNI FNVSNKPNDAVECVGKKRSREDDLFSLGCSKKQAIINI
L*

>G681 (1..804)

ATGGGGAGGACGACATGGTTCGACGTCGACGGGATGAAGAAAGGAGAGTGACGGCAGAG
GAAGACCAGAAGCTCGGCGCTTACATCAACGAGCATGGCGTTTGTGATTGGCGTTCCCTC
CCCAAAAGAGCTGGTTTGCAGAGATGTGGAAGAGCTGCAGATTAAGGTGGCTTAATCTAT
CTAAAGCCTGGGATTAGAAGAGGCAAATTCCTCCTCAAGAAGAAGAAGAAATCATCCAA
CTTCATGCTGTTCTCGGAAACAGGTGGGCAGCCATGGCGAAGAAGATGCAGAATCGAACA
GACAATGATATCAAGAACCATTGGAACCTTGTCTCAAGAAAAGACTTTCGAGAAAGGGA
ATCGACCCTATGACCCACGAGCCCATCATCAAACACCTCACCCTCAATACCACTAACGCA
GATTGTGGTAACCTTCCACCACGACGTCCCCGTCGACGACGGAAGCTCTCCTTCTCTC
GGCTCGTCTCGTCTTCTTAACAACTCGCCGAGGTATCTCATCTAGACAACATAGTCTC
GATAGGATCAAGTACATCTTGTGCAATTCAATAATCGAAAGCAGTGATCAAGCAAAGAG
GAAGAAGAAAAGAAGAAGAAGAAGAAGAGATTCAATGATGGGTGAGAAGATTGAC
GGTAGTGAAGGAGAAGATATTGAGATTTGGGGCAGGAGGAAGTTAGGCGTTTAATGGAG
ATTGATGCAATGGATATGTACGAGATGACTTCGTACGACGCTGTCATGTACGAGAGTAGT
CACATACTTGATCATCTCTTTTACTTAATATAGTGTGACTGTGTGAGTGCATGCATGTT
>G681 Amino Acid Sequence (domain in AA coordinates: 14-120)

MGRTTWFDVDMKKGEWTAEDQKLGAAYINEHGVCDWRSPLPKRAGLQRCGKSCRLRWLNY
LKPGIRRGKFTPQEEEBIIQLHAVLGNRWAAAMAKKMQNRDNDIKNHWNCLKKRLSRKG
IDPMTHPIIKHLTVNTNADCGNSSTTTSPSTTESSPSSGSSRLNKLAAGISSRQHS
DRIKYLNSNIISSDQAKEEEEEEEEEERDSMMGQKIDGSEGEDIQIWGEEBVRRLME

IDAMDMYEMTSYDAVMYESSHILDHLF*

>G1482 (1..996)

ATGAAGATCAGGTGCGACGTCTGCGATAAAGAAGAAGCGTCGGTGTGTTTGCACGGCCGAC
GAAGCATCTCTCTGCGGCGGCTGCGACCACCAAGTCCACCACGCTAACAAACTCGCCTCT
AAACATCTCCGTTTCTCTCTCTTTATCCTTCTTCTTCCAACACCTCTCTCTCTCTG
GACATCTGTGAGGATAAAAAAGCTCTGTTGTTCTGTCAACAAGATAGAGCTATTTTATGC
AAAGATTGCGATTTCATCGATCCACGCTGCGAACGAACACACAAAGAAACACGATAGGTTT
CTTCTTACAGGGGTTAAGCTCTCTGCAACATCGTCTGTTTACAAACCTACTTCGAAATCT
TCTTCTTCTTCTTCAAGCAACCAAGATTTCTCTGTCCTGGATCATCAATCTCTAATCCT
CCTCCTCTCAAGAAACCTCTCTCAGCTCCTCCTCAGAGCAACAAGATCCAACCCCTTTCG
AAGATCAACGGCGGTGATGCGTCGGTGAATCAGTGGGGATCCACAAGCAGATTTCTGAG
TATTTGATGGATACGTTACCTGGTTGGCACGTTGAGGATTTCTCTGATTCCTCTCTTCTCT
ACTTATGGTTTCTCTAAGAGTGGTGTGATGATGGAGTGTACCATATATGGAACCAGAA
GATGACAACAACAACAAACAACAACAACAACAACAACAACAACAATACAGTG
TCACTTCCATCTAAGAATTTAGGGATTTGGGTCCCTCAGATTCACAAACTCTTCTCTTCT
TCATACCCAAATCAATACTTTTCTCAAGACAACAACATACAGTTTGGGATGTACAAACAA
GAAACATCACCAGAAGTAGTGTCTTTTGTCTCAATAACAAACATGAAACAACAAGGACAG
AACAACAAGAGATGGTATGATGATGGTGGCTTCACTGTCCACAGATCACTCCTCCTCCT
CTTTCTCTAATAAAAAGTTTAGATCTTCTGGTAA

>G1482 Amino Acid Sequence (domain in aa coordinates: 5-63)
MKIRCDVCDKEEASVFCFADEASLCGGCDHQVHHANKLASKHLRFSLLYPSSSNTSSPLC
DICQDKKALLFCQODRAILCKDCDSSIHAANEHTKKHDFLLTGVLKLSATSSSVYKPTSKS
SSSSSSNQDFSVPGSSISNPPLKKPLSAPPQSNKIQPFKINGGDASVNQWGSTSTISE
YLMDTLPGWHVEDFLDSSLPTYGFSKSGDDDGVLPMPEPEDDNTKRNNNNNNNNNNTV
SLPSKNLGIWVPQIPQLPSSYPNQYFSQDNNIQFGMYNKETSPVVVSFAPIQNMKQGGQ
NNKRWYDDGGFTVPQITPPPLSSNKKFRSFW*

>G225 (157..441)

CTCTCTCTCTCACTCTTTCTTTTCCGAGAACCACAAAAAAGCTACTATTAATCC
TTCCCTCGTGAGGAAATCATTTCTTCTGTTTCTCGAGATTATTCTCTTCTCTCTCT
CTTTCTCTGTGTGTTTCGTGTCTTCAGATTAGTTCGATGTTCGTTTCAGACAAGCGGAA
AAAATGGATAAACGACGACGAGACAGAGCAAAGCCAAGGCTTCTTGTTCGGAAGAGGTG
AGTAGTATCGAATGGGAAGCTGTGAAGATGTCAGAAGAAGAAGATCTCATTTCTCGG
ATGTATAAACTCGTTGGCGACAGGTGGGAGTTGATCGCCGGAAGGATCCCGGACGAGCG
CCGGAGGAGATAGAGAGATATTGGCTTATGAAACACGGCGTCGTTTTTGGCAACAGACGA
AGAGACTTTTTTAGGAAATGATTTTTTTGTTTGGATTAAAGAAAATTTCTCTCTCT
AATTCACAAGACAAGAAAAAAGGAAATGTACCTGTCTTGAATTACTATTTTGAATGT
ATAATATCTATATATATAAGAAGAAAAAATTGCTTAGGAATTT

>G225 Amino Acid Sequence (domain in AA coordinates: 39-76)
MFRSDKAEMDKRRRRQSKAKASCSEEVSSIWEAVKMSSEEBDLISRMVYKLVGDRWELI
AGRIPGRTPPEIERIERYWLMKHGVVFANRRRDFFRK*

>G226 (10..348)

CCAGTAGTTATGGATAATACCAACCGTCTTCGTCTTCGTGCGGTCCAGTCTTAGGCAA
ACTAAGTTCACTCGATCCCGATATGACTCTGAAGAAGTGAGTAGCATCGAATGGGAGTTT
ATCAGTATGACCGAACAAGAAGAAGATCTCATCTCTCGAATGTACAGACTTGTGCGTAAT
AGGTGGGATTTAATAGCAGGAAGAGTCGTAGGAAGAAAGGCAAATGAGATTGAGAGATAC
TGGATTATGAGAACTCTGACTATTTTTCTCAAAACGACGACGTCTTAATAATTCTCCC
TTTTTTCTACTTCTCTCTTAATCTCCAAGAAAATCTAAAATTGTAAAGAAATCAAAT
AAAAGCTTTCAATCATAAAAAGTAGAACAAATCTTGAATGTCTTCTCA

>G226 Amino Acid Sequence (domain in AA coordinates: 28-78)
MDNTNRLRLRRGPSLRQTKFTRSRDYSEEVSSIWEFISMTEQEEDLISRMVRLVGNRWD
LIAGRVVGRKANEIERIERYWIMRNSDYFSHKRRRLNNSPFFSTSPNLQENLKL*

>G9 (81..1139)

GTGTTTCTTCTTCTGCTAAAAGGTTATAATTTTTGTTTCTTGGTTTGGTGAGAATCTTC
AAGAACTGAAACAAAGAAAATGGATTCTAGTTGCATAGACGAGATAAGTTCTCCACTT
CAGAATCTTTCTCGCCACCACCGCCAAGAAGCTCTCTCCTCCTCCGCGGCGGCTTAC
GCCTCTACCGGATGGGAAGCGGCGGAGCAGCGTCGTGTTGGATCCCGAGAACGCGCTAG
AGACGGAGTCACGAAAGCTACCATCTTCAAAATACAAAGGTGTTGTTCTCAGCCTAACG

GAAGATGGGGAGCTCAGATCTACGAGAAGCACCAACGAGTATGGCTCGGGACTTTCAACG
AGCAAGAAGAAGCTGCTCGTTCTTACGACATCGCAGCTTGATAGATTCCGTGGCCGCGACG
CCGTCGTCAACTTCAAGAAGCTTCTGGAAGACGGCGATTTAGCTTTTCTTGAAGCTCACT
CAAAGGCCGAGATCGTCGACATGTTGAGAAAAACACTTACGCCGACGAGCTTGAACAGA
ACAATAAACCGGCAGTTGTTTCTCTCCGTGACGCTAACGGAAAACGTAACGGATCGAGTA
CTACTCAAACGACAAAGTTTTAAAGACGTGTGAAGTTCTTTTCGAGAAGGCTGTTACAC
CTAGCGACGTTGGGAAGCTAAACCGTCTCGTGATACCTAAACAACACGCCGAGAAACACT
TTCCGTTACCGTCACCGTCACCGGCAGTGACTAAAGGAGTTTGTATCAACTTCGAAGACG
TTAACGGTAAAGTGTGGAGGTTCCGTTACTCATACTGGAACAGTAGTCAAAGTTACGTGT
TGACCAAGGGATGGAGTCGATTTCGTCAAGGAGAAGAATCTTCGAGCCGGTGATGTTGTTA
CTTTCGAGAGATCGACCGGACTAGAGCGGCAGTTATATATTGATTGGAAGTTTCGGTCTG
GTCCGAGAGAAAAACCGGTTACGGTGGTGGTTCGGCTTTTCGGAGTTGATATCTTTAATG
TGACCACCGTGAAGCCAAACGACGTCGTGGCCGTTTTCGGTGGAAAGAGATCTCGAGATG
TTGATGATATGTTTGCCTTACGGTGTTCGAAGAAGCAGGCGATAATCAATGCTTTGTGAC
ATATTTCTTTTCCGATTTTATGCTTTTCGTTTTTAAATTTTTTTTTTGTCAAGTTGTGT
AGGTTGTGATTATGCTAGGTTGTATTTAGGAAAAGAGATAAGACC

>G9 Amino Acid Sequence (domain in AA coordinates: 62-127)

MDSSCIDIEISSSTSEFSATTAKKLSPPPAALRLYRMGSGSSVLDPENGLTESRKL
PSSKYKGVVPQPNRWGAQIYEKHQVRVWLGTFNEQEEAARSYDIAACRFRGRDAVVNFKN
VLEDGDLAFLEAHSKAEIVMLRKHTYADELEQNNKRLFLSVDANGKRNGSSTQNDKV
LKTCEVLFEKAVTPSDVGLNRLVLPKQHAKEHFLPLSPSPAVTKGVLINFEDVNGKVWR
FRYSYWNSSQSYVLTKGWSRFVKEKNLRAGDVVTFERSTGLERQLYIDWKVRSRGPENPV
QVVVRLFGVDIFNVTTVPKNDVVAVCGGKRSRDVDDMFALRCSKKQAIINAL*

>G1040 (51..863)

CTTTGATCTCCACTATTTAAGTAGACAAGAATCATAAAGAAAATAGTGAGATGATGATGT
TAGAGTCAAGAAACAGTATGAGAGCTTCAAACCTCAGTCCCAGATCTGTCTCTTCAGATCA
GTCTTCTTAATATACGCGCGGAAAACCTCTTACGGCGGTGACCGGAGCTCCACAAGCA
GTGATTCTGGAAGCAGCCTCAGTGACCTGAGCCATGAGAACAATCTTCAACAAACCTC
TCTTGAGCTTAGGATTTGACCATCATCATCAAAGGCGCTCAAACATGTTCCAACCTCAA
TCTACGGTCGAGATTTCAAGAGAAGCTCATCATCAATGGTTGGTCTTAAACGAAGCATT
GTGCTCCAAGAATGAGATGGACTTCTACTCTTCATGCTCACTTCGTCCATGCTGTTCAAC
TTCTTGCGCGCCATGAAAGAGCAACGCCTAAATCAGTGTTGGAGCTCATGAATGTGAAGG
ATCTAACCTTAGCTCATGTCAAGAGTCACTTGACAGATGTATAGAACAGTGAAATGCACTG
ATAAAGGATCACCAGGAGAAGGAAAGGTAGAGAAAGAGGCAGAGCAGAGGATAGAGGACA
ATAATAAATGAAGAAGCTGATGAAGGAAGTACACAAATTCGCCAAACTCATCATCTG
TGCAAAAGACCCAAAGAGCTTATGTTTCATCGACAAAGGAAGTATCTAGGAGCATATCTA
CACAAGCATATTTCACTTGGGAACAACCTCATCACACTAAGGCCAATGAAGAGAAAGAGG
ATACCAACATTCATCTCAATTTGGATTTACATTGGGCGGCCTAGTTGGGGGATGGAATA
TGCGGAACCCCTCCAGTGATTTAAACCCTTCTCAAGTGCTAATTGCCTTAAGCTACAACAAA
TAAGTCAGCTTAGGTTACAGTTTAAACATAATTTAACTTGTGTTTGTATCATATGAGCTT
CGGAAGAATCATATTATCATCATATATGAACCTTCTTCCAAGAATGTTCTATGAGTTTTT
TGATATGTATAATCAAGAGAATCGTTTGAAGTAAAAA

>G1040 Amino Acid Sequence (domain in AA coordinates: 109-158)

MMLESRNSMRASNSVPDLSLQISLPNYHAGKPLHGGDRSSTSSDSGSSLSDLSHENNFF
NKPLLSLGFDDHHHQRRSNMFQPIYGRDFKRSSSMVGLKRSIRAPRMRWTSTLHAHFVH
AVQLLGGHERATPKSVLELMNVKDLTLAHVKSHLQMYRTVKCTDKGSPGEGKVEKEAEQR
IEDNNNNEEADGTDNTPNSSSSVQKTQRASWSSTKEVRSISTQAYSHLGTTHHTKANE
EKEDTNIHLNLDFTLGGGLVGGWNMRNPPVI*

>G2114 (64..1311)

ATAAAACGAAACCTTATACATATAAACTAAGAGCGAGAAAGACAGCTAGAGAGAGAGAGA
GAGATGAAGAAATGGTTGGGATTTTCATTGACACCTCCTTTGAGAATCTGCAATAGTGAA
GAAGAAGAACTTAGGCATGACGGTTCCGATGTTTGGAGATATGATATTAACCTTGTATCAT
CATCATCATGATGAAGACGTTCCAAAGGTGGAAGATCTCCTCTCAAACCTCTCATCAAACC
GAGTATCCTATAAACCATAACCAAAACCAATGTCAACTGCACCACTGTGGTTAACAGGTTA
AACCACCCGGTTACCTTCTCCACGACCAACCGTAGTTACACCACATTACCCGAACCTA
GATCCGAACCTTAGCAATGATTATGGAGGTTTTGAGAGGGTCGGTTCGGTCTCGGTTTTT
AAATCTTGTTAGAGCAAGGCACTCCAGCATTCCCACTCTCGAGTCATTACGTTACTGAA

GAGGCTGGTACGAGCAATAATATTAGTCATTTTAGTAACGAAGAGACTGGTTATAACACC
AATGGCTCAATGCTATCATTGGCTTTGAGCCATGGGGCTTGTTCTGATTTGATCAACGAA
TCGAATGTATCCGCACGGGTGGAAGAACCAGTTAAGGTAGATGAGAAGCGGAAGAGATTG
GTTGTTAAACCTCAGGTAAAGGAATCCGTTCCCTCGGAAGTCGGTTGATAGTTATGGACAA
AGAACTTCTCAGTATCGTGGAGTTACAAGGCATAGATGGACAGGGAGATATGAAGCTCAC
TTATGGGATAATAGCTGTAAGAAGGAGGGACAGACAAGGAGAGGAAGACAAGTGTATCTT
GGAGGGTATGATGAGGAGGAGAAAGCAGCGAGGGCATATGATTTAGCGGGCTCTGAAGTAT
TGGGGTCCCTACCACTCACTTAAATTTCCCTTTGAGTAATTACGAAAAGGAGATCGAGGAA
CTCAATAACATGAATCGGCAAGAATTTGTTGCCATGTTGAGGAGGAATAGCAGCGGGTTT
TCGAGGGGAGCTTCCGTGTATAGAGGAGTTACAAGGCATCATCAACATGGAAGGTGGCAA
GCCAGAAATTGGAAGAGTTGCTGGAACAAGGACTTGTACCTTGAACATTTAGCACGCAA
GAAGAAGCAGCGGAGGCGTACGATATCGCGGCAATTAAATTCAGAGGCCTAAACGCTGTA
ACCAATTTTCGATATAAATAGATATGACGTGAAGAGGATATGTTCAAGCTCAACGATTGTT
GATAGCGACCGAGGCCAAACATTTCCCACAGCTCTGGCGCCGCCACTAACCGACACCG
TAAACTCCTCGCCGGAGAGACTATTTCCACGTACGGTTGGTTTGGAGGAAATAAGTTCGTC
CAGTCTGTTTAAATCATTTATGGTTTAAATAACATATATTCCTAAGTAATTGAGGCCGGTC
TACATATATACAACTTTTTTAGCAAATTAAGTTATCAGAATCCACTATATATTATTCTCT
>G2114 Amino Acid Sequence (conserved domain in AA coordinates: 221-297, 323-393)
MKKWLGLFSLTPPLRICNSEEELRHGSDVWRYDINFDDHHHDEDPKVEDLLSNHQTE
YPINHNQTNVNCCTTVNRLNPPGYLLHDQTVVTPHYPNLDPNLSNDYGGFERVGSVSVFK
SWLEQGTAPFPLSSHVYTEEAGTSNNISHFSNEETGYNTNGSMLSLLSLSHGACSDLINES
NVSARVEEPVKVDEKRRKRLVVKPQVKESVPRKSVDSYQRTSQYRGVTRHRWTGRYEHL
WDNSCKKEGQTRRRGRQVYLLGGYDEEKAARAYDLAALKYWGPTTHLNFPLSNYEKEIEEL
NNMNRQEFVAMLRNSSGFSRGASVYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTSTQE
EAAEAYDIAAIKFRGLNAVTFNFDINRYDVKRICSSSTIVSDQAKHSPTSSGAGH*
>G450 (65..751)
GAGTTATCGAGAGAGAGAGAAAACATATTCTGATTTAAGACATATATAGACAGCAAGAAG
AGATATGAACCTTAAGGAGACGGAGCTTTGTCTTGGCCTCCCCGGAGGCCTGAAACCGT
TGAAAGTCCGGCCAAGTCGGGTGTTGGGAACAAGAGAGGCTTCTCCGAGACCGTTGATCT
CAAACCTTAATCTTCAATCTAACAACAAGGACATGTGGATCTCAACACTAATGGAGCTCC
CAAGGAGAAGACCTTCCCTTAAGACCTTCTAAGCCTCCTGCTAAAGCACAAAGTGGTGGG
TTGGCCACCGGTGAGGAACCTACCGGAAAAATGTTATGGCTAATCAGAAGAGCGGCGAAGC
AGAGGAGGCAATGAGTAGTGGTGGAGGAACCGTCGCTTTGTGAAGGTTTCCATGGATGG
AGCTCCTTATCTTTCGGAAGGTTGACCTCAAGATGTACACCAGCTACAAGGATCTCTCTGA
TGCCTTGGCCAAAATGTTTCACTCCTTTACCATGGGGAGTTATGGAGCACAAAGGGATGAT
AGATTTTCATGAACGAGAGTAAAGTGTGATGATCTGTTGAACAGTTCTGAGTATGTTCCAAG
CTACGAGGACAAAGATGGTGAAGTGGATGCTCGTTGGTGTATGTCCTGGCCGATGTTTGT
CGAGTCATGCAAAACGTTTTCGCGCATAATGAAAGGATCCGAAGCAATTGGACTTGCTCCAAG
AGCAATGGAGAAGTTCAAGAACAGATCATGAACAAAAAAGAGGACAAATATGCATTG
ATTTTTTTTTTTTTTGGTATTGTTATGATCATGTGTTTTAATTAAATATAGGAAGGATA
TAGGAAAATATAATTGTTTACAAAAAATAACTTTAAATATGCTTTTTTTTTTTTTTGA
AATTAGTCTGTGTTTTGTTTTTCATCTCTTAATTAGTAGAAATCATTTTTTTAATATGTAA
TTGTGATAGTAAATCTATAGAGTTCGTA
>G450 Amino Acid Sequence (domain in AA coordinates: TBD)
MNLKETELCLGLPGGTETVESPAKSGVGNKRGFSETVDLKLNLQSNKQGHVDLNTNGAPK
EKTFLKDPKPPAKAQVVGWPPVRNYRKNVMAKQSGEABEAMSSGGTVAFVKVSMGDA
PYLRKVDLKMYSYKDLSDALAKMFSSFTMGSYGAQGMIDFMNESKVMDDLNSSEYVPSY
EDKDGWMLVGDVPMFVESCRLRIMKGSEAIGLAPRAMEKFKNRS*
>G584 (40..1809)
AAAAAGTCTTCTCTTTTATAACTACGTACAGAGAACTGTTATGTCTCCGACGAATGTTCAA
GTAACCGATTACCATCTCAACCAATCAAAAACGGATACAACAAATCTCTGGTCAACCGAC
GACGATGCATCGGTAATGGAAGCTTTTCATCGGCGGCGGCTCCGATCATTTCTCTTTTT
CCTCCACTTCTCCTCCTCCTTCTCAAGTCAACGAAGATAATCTCCAGCAACGCTCTC
CAAGCTTTAATCGAAGGAGCAACGAGAACTGGACTTACGCGGTGTTCTGGCAATCATCT
CACGGTTTCGCGCGAGAAGACAACAACAACAACACAGTGTGTTAGGTTGGGGAGAT
GGTTATTACAAAGGAGAAGAAGAGAAGTCTAGAAAGAAGAAATCCAGCTAGTGCA
GCTGAACAAGAGCATCGTAAGAGAGTATAGAGAGCTCAACTCTTAATCTCCGGTGGT

GTAGGAGGAGGAGATGAAGCTGGAGATGAAGAAGTTACAGATACTGAATGGTTCTTCTTA
GTTTCAATGACACAGAGCTTTGTCAAGGGTACTGGTTTACCTGGTCAAGCTTTCTCAAAT
TCAGACACGATTGGTTATCTGGTTCTAATGCTTTAGCTGGATCAAGTTGTGAGAGAGCT
CGTCAAGGTCAGATTTATGGGTTACAAACAATGGTGTGTAGCGACAGAGAATGGTGTG
GTTGAGCTTGGTTTCGTCGGGAGATTATTCATCAAAGTTCAGATCTTGTGATAAAGTTGAC
ACCTTTTTCAATTTTAAACAATGGTGGTGGTGAATTTGGTTCTTGGGCGTTAATTTGAAT
CCAGATCAAGGAGAGAATGATCCAGGTTTGTGGATTAGTGAACCTAATGGTGTGACTCT
GGTCTTGTAGCTGCTCCGGTGATGAATAATGGTGGAAATGACTCAACTTCTAATTCTGAT
TCTCAACCAATTTCTAAGCTTTGTAAATGGAAGCTCTGTTGAAAACCTAACCTTAAAGTT
CTGAAATCTTGTGAAATGGTGAATTTCAAGAATGGGATTGAGAATGGTCAAGAAGAAGAT
AGTAGTAATAAGAAGAGATCACCAGGTTTTCGAATAATGAAGAAGGGATGCTTTCTTTTACC
TCTGTTCTTCCATGTGACTCGAATCACTCTGATCTTGAAGCTTCAGTGGCTAAAGAAGCT
GAGAGTAACAGAGTTGTGGTTGAACCGGAGAAGAAACCGAGGAAACGAGGGAGAAAACCG
GCGAATGGAAGAGAGAGAGCCTTTGAATCATGTAGAGGCAGAGAGACAGAGAAGAGAGAAG
TTGAATCAGAGATTCTATTCTTTAAGAGCTGTGGTTCCTAATGTGTCTAAGATGGATAAA
GCTTCTCTATTAGGAGATGCTATTTTCGTATATCAGTGAGCTTAAGTCTAAGTTGCAAAAG
GCTGAATCTGATAAAGAGAGTTGCAGAAGCAGATTGATGTGATGAATAAAGAAGCGGGA
AATGCGAAAAGTTCCGTAAAAGATCGAAAATGTTGAATCAAGAATCAGAGTGTGTGATA
GAGATGGAGGTTGATGTGAAGATTATTGGTTGGGATGCAATGATAAGGATTCAATGTAGT
AAGAGGAATCATCCTGGTGCTAAGTTCATGGAAGCACTTAAGGAGTTGGATTGGGAAGTG
AATCATGCGAGTTTATCGGTAGTGAATGATCTTATGATCCAACAAGCGACTGTGAAAATG
GGGAATCAGTTTTTTCACGCAAGATCAACTCAAGGTTGCTCTAACGGAGAAAGTTGGAGAA
TGTCATGAATTGAAGTCAGCATCTTTAGGGCTAATACACCGGAGAATACTGCGAAAAGT
CGAAAACAACGATCATAGTATAAGCCGCGTAAAAAGTGTAAACCTTTTACACAAGTTT
CTCTAGTGAATGTAGTTGTAACTCTATTGTGTAAAGGTAATTTGTAGTACCCACTTGT
TGCTATTGAATGCTTGTGTAGAGAGGATTCTTAGTGTAGTATATGATTAGGTTGGGGTTG
TTGTTTCATGAGATAAATAAATGTGTTGATCAATGGTTAAGTCTTTGGTTTGTGGTGT
ATGTATGTAATAAGGCTTTTGTGTAGAAATAAGACAAATGGGACTGAAGTTGGAGTTTAA
AA

>G584 Amino Acid Sequence (domain in AA coordinates: 401-494)
MSPTNVQVTDYHLNQSKDITNLWSTDDASVMEAFIGGGSDHSLFPPLPPPPLPQVNE
DNLQORLQALIEGANENWTYAVFWQSSHGFAGEDNNNNNTVLLGWGDGYKKEEEKSRKK
KSNPASAAEQEHRKRVIRELNSLISGGVGGGDEAGDEEVDTEWFFLVSMQTQSFVKGTGL
PGQAFNSDITWLSSNALAGSSCERARQGQIYGLQTMVCVATENGVELGSSEI IHQSS
DLVDKVDTFNFNNGGEGFSWAFNLNPDQGENDPGLWISPNGVDSGLVAAPVMNNGGN
DSTNSDSQPISKLCNGSSVENPNPKVLKSCMVNFKNGIENGQEEDSSNKKRSPVSNNE
EGMLSFTSVLPDCSNHSDLEASVAKEAESNRVVVEPEKKPRKRGRKPANGREEPLNHVEA
ERQRREKLNQRFYSLRAVVPNVSKMDKASLLGDAISYISELKSLQKAESDKEELQKQID
VMNKEAGNAKSSVKDRKCLNQESSVLIEMEVDVKIIGWDAMIRIQCKSRNHPGAKFMEAL
KELDLEVNHASLSVNDLMIQQATVKMGNOFFTQDQLKVALTEKVGECP*

>G668 (1..1056)

ATGGGAAGACCACCTTGCTGTGAAAAGATTGGAGTGAAGAAAGGGCCATGGACACCAGAG
GAAGACATCATCTTGGTTTCTTACATCCAAGAACATGGTCCTGGAACTGGAGATCTGTC
CCAACACACACAGGTTTAAAGATGTAGCAAGAGCTGCAGATTGAGATGGACTAATTATCTT
CGACCCGGTATTAAGCGTGGAAATTTTACTGAGCATGAAGAGAAGACAATTGTTTCATCTT
CAAGCCCTTTTAGGCAACAGATGGGCAGCCATAGCATCATACCTTCCAGAAAGGACAGAC
AATGATATAAAGAATAATTGGAACACTCACTTGAAGAAGAAGCTCAAAAAGATTAAATGAA
TCTGGTGAAGAAGATAATGATGGTGTCTCTTCATCAAACACTAGTTCAAAAAGAACCAT
CAAAGCACTAACAAAGGTCAATGGGAAAGAAGACTTCAGACAGACATTAACATGGCAAAA
CAAGCTCTTTGTGAGGCCTTGTCTTTAGACAAACATCATCCACTCTTTCATCATCTTCA
TCATTACCGACACCAGTAATCACACAACAAAACATCCGTAACTTCTCATCAGCTTTGCTT
GACCGTTGTTATGATCCATCCTCTTCTTTCATCTACCACAACCACCACTACAAGCAAC
ACTACTAATCCATACCCATCAGGGGTATATGCGTCAAGTGTGAGAACATCGCCCGGTTG
CTTCAAGATTTTCATGAAAGACACACCAAGGCTTTAACTTTATCATCTTCATCTCCGGTT
TCAGAGACTGGACCACTCACTGCTGCAGTCTCGGAAGAAGGTGGAGAAGGGTTTGAACAA
TCTTCTTTCAGCTTCAATTCAATGGACGAAACTCAAACTTGACTCAGGAGACAAGCTTC
TTCCATGATCAAGTGATCAACCCGGAAATAACAATGGACCAAGATCATGGTCTAATATCA

GACGTGTTCAACAAGGATCCATATGTGAAGTTCATGCTGAGATCCCTAGCTTCGTGATC
GTTAAACCACGAACAAAGGCTTGTGGTAAAACCGATGGATGTGATTCGGGTTGCTGGAGG
ATCATTGGTCGTGATAAGCTGATAAAGTCGGAGGAGACTGGTAAGATTCTAGGGTTCAAG
AAGATACTCAAGTCTCTGCTAAAGTGGAAACCTAGAGAATACAAGAGAAGTTTGGTAATG
GAAGAGTATAGGCTTACCAATAACTTCAACTGGAAGCAAGATCATGTGATTTGCAAGATT
CGGCTTTTGTGTTGAAGCAGAAATTAGTTTCTTGCTAGCCAAGCATTCTACACTACATCA
GACTCACTTCCTCGAAATGTGCTGTTGCCAGCTTATGGATTCTGTTACCAGATAAAACAA
GAGGAGGACGAATTTTATCCGGTGACGATAATGATTTTCAAGGAAAAGATTGGCCTAGC
TACGTTACCAACAACGTGTATTGTCTGCATCCATCGGAGCTTGTGAATGTTACGATGGG
AAGTTTCATGATAACGGAATCTGCATCTTCGCTAACAGGACTTGTGGTGTAACCGATAAA
TGCAATGAAGGTTACTGGAAGATTAAGCACCGTGAGAAGCTGATCATGTACGGTACGGG
CAGACCATTGGTTGGAAGAAAGTTTTTCAGTTTTATGAAACGGAGAAAAGAACATTTT
GGTAATGGAGAAGAAGTGAAGGTAACCTGGACTCTAAAAGAGTATAGGCTTACCAGAAAA
ATGAACAAGATAAAGTGGTGTGCGTTATCAAGTATAAGGTAAAGTGTTCACCGAGGATA
ACTAGCTAGGAGCTTCTACTCTTGGTTTCATGATCGATGCGACCGCTCTAGACAGGCCTC
GTACCGGATCCTCTAGCTAGAGCTTTCGTTTCGTATCATCGGTTTCGACAACGTTTCGTCA
>G1463 Amino Acid Sequence (conserved domain in AA coordinates:9-156)
MRFFSLVPLFLGRFSFVMDTLNNAEDEVIIISRYLKPMIVNRVSWPDLFIEDADVFNKD
PYVKFHAIEPSFVIVKPRTKACGKTDGCDSGCWRIIGRDKLIKSEETGKILGFKKILKFC
LKWKPREYKRSVMEERYLTNNFNWKQDHVICKIRLLFEAEISFLLAKHFYTTSDSLPRN
VLLPAYGFCSPDKQEEDEFYPVTIMISEGKDWPSYVTNNVYCLHPSELVNVHDGKFHDNG
ICIFANRTCGVTDKCNENYWKIKHREKLIMSRYGQTIGWKVQFYETEKERHFNGGEEV
KVTWTLKEYRLTRKMNKNKVVCVIKYKVKCLPRITS*
>G1944 (236..1306)
TCGACCTTCTAATTTCCAACCTCTGTTCTTAGCAATATATTTTTTCTCCAAAAATAATT
CTCAGTTTGATTTTCTTCTCTAGCTCTTAAGTATATTTCTTTGTTGTTATTTATCTTTT
AATCCTTTAATCTCATCTTTGTTTATCTTTAATCAAAACCCAAAATTTACATGGGTTCTT
GAAAATCTAGAGAAATAAAGGAAACATAACAAAAATAGAAAGAAAAAGAAGCTAATGGT
CTTAAATATGGAGTCTACCGGAGAAGCTGTTAGATCAACCACCGGTACGACGGTGGTAT
TACGGTGGTTAGATCCGACGCGCGGTACAGATTTCCACGTAGCTCAAAGATCAGAAAGCTC
AAACCAATCTCCACCTCTGTCACTCCTCCTCCACCACAGCCATCGTCTCATCACACAGC
TCTCCGCGCGTGCAAATTTGACGCGTGACGACTACGACTACGACGCGCGCGATGGAAGG
TATCTCCGGTGGACTGATGAAGAAGAAGCGTGACGCGCAAGGAAGTATGGACCGGACGG
GACTGTTGTAGCGTTATCTCTAAACCGATTTCATCAGCGCGCGCGCGTCCGATCTTCC
GCCGCGGAGTTACACGTCATCGATTTCTCCGCTTCTGAGAAACGTAGCAAAGTGAAACC
AACGAACTCGTTTAAACAGAACAAAGTATCATCACCAAGTTGAGAATTTGGGTGAATGGGC
TCCTTGCTCCGTCGGTGGTAATTTACACCTCATATAATCACAGTCAACACCGGCGAGGA
TGTAACAATCGATGAATAATCTCGTTTTTCGCAACAAGGACCTCGCTCTATTGTGTTCTGTC
AGCAAACGGTGTATTTCAGCGTTACACTTCGTAGCCAGATTCTCTGGCGGCACATT
GACATACGAAGTCCGTTTGAATATTATCATTATCCGGGTCATTATGCCTAATGATTC
AGGCGGAACACGAAGTAGAACGGGAGGAATGAGTGATCGTTAGCAAGTCCCGATGGACG
TGTAGTAGGCGGTGGCCTCGCCGGTTTACTAGTAGCCGCGAGTCCGGTTCAGGTGGTGT
AGGAAGTTTTTTAGCGGGCACTGACCATCAAGATCAGAAACCGAAAAAGAACAAACATGA
TTTCATGTTGTCGAGTCTTACCGTGCATTCCTATCTCTAGTGACGCTGATCACCGGAC
AATCCATTCCGTCTCGTCTCTTCCGGTCAATAATAATACATGGCAGACTTCTTTAGCTTC
CGATCCAAGAACAAAGCATACCGATATTAATGTCAATGTAACTGAAATCCAATCTTTCT
CTGTATTTTCTGTTAACAAGTTTGATTTGGTTGTTTATCTACATTAGGATTTTACTAAAA
TGGTAGTATTATTATAGGTTTTAGGGTCTTTATTTTGGTCCACTGTTGTCACTTGTA
GGATA
>G1944 Amino Acid Sequence (domain in AA coordinates:87-100)
MVLNMESTGEAVRSTGNDGGITVVRSDAPSDFHVAQRSESSNQSPSTVTPPPQPSHH
TAPPLQISTVTTTTTTTAAAMEGISGGLMKKRGPRKYGPDGTVVALSPKISSAPAPSH
LPPSSHVIDFSASEKRSVKPTNSFNRTKYHHQVENLGEWAPCSVGGNFTPHIITVNTG
EDVTMKIISFSQQGPRISCVLSANGVISSVTLRQPDSSGGTLTYEGRFEILSLSGSFM
DSGGRSRTGMSVSLASPDGRVVGGGLAGLLVAASPVQVVVGSFLAGTDHQDQKPKKNK
HDFMLSSPTAAIPISSAADHRTIHSVSSLPVNNNTWQTSLASDPRNKHTDINNVNT*
>G2383 (37..990)

GACCTCTTTGATCCCTTCATTCCCATCAAACAACCATGTTTCCTTCTTTTCACTACTCAC
ATTCAAAGCCCTAATTCTCACCATCACTACTCTTCGCCTTCTTTTCCCTTCTCTTCCGAT
TTTCTTGAGAGTTTGTGATGAATCCTTCTTGATAAACCAATTCTTGTTACAGCAGCAAGAT
GTAGCAGCAAAATGTTGTTGAATCTCCTTGGAATTTTGCAGAAGCTTGAGCTTAAGAAG
AAGAATGAGAAGTGTGTTGATGGAAGCACCTCACAAGAGGTTCAATGGAGAAGGACGGTC
AAAAAAGGGACAGGCATAGTAAGATCTGCACGGCTCAAGGTCCTAGAGACCGGAGGATG
AGGCTGTCTCTCAGATTGCTCGCAAGTTTTTCGATCTTCAAGACATGTTGGGTTTCGAC
AAGGCGAGCAAGACGATTGAATGGCTTTTCTCCAAATCAAAGACTTCCATCAAACAACCTT
AAAGAAAGAGTGGCTGCATCGGAAGGAGGAAAGGATGAACATCTCCAGGTTGATGAA
AAGGAAAAGGATGAGACACTGAAGTTGAGAGTCTCAAAGAGAAGAACAAGACTATGGAG
AGCTCTTTTAAAGACTAAAGAGTCGAGAGAGAGAGCTAGAAAGCGAGCAAGAGAGAGAACA
ATGGCAAAGATGAAGATGAGATTATTTGAGACCTCGGAAACAATTTAGATCCTCATCAA
GAACTAGAGAGATCAAGATAACCAATGGTGTAACAATTACTAGAAAAGGAAAATAAGAA
CAAGAATGGAGTAATACTAATGATGTTTACATGGTAGAGTATCAAATGGATTCTGTGAGC
ATCATAGAGAAAGTTTCTTGGACTAACCAGTACTCTAGCTCCTCTTCCATTTTGGTGAC
TCCGAGGAATGTTACACAAGTCTTAGTTAGTAAGAGGTACAATTTAGCAGCAGGTAAC
AGCAATGTGTTAACTAAAAACCCTAATTGAGTAATGCAGTTTTGATTAAATATTAGCTTTT
TGGTAATTCAGGAATGTCGACACCAAGGG
>G2383 Amino Acid Sequence (conserved domain in AA coordinates:89-149)
MFPSFITHIQSPNSHHYSSPSFPFSSDFLESFDESFLINQFLQQQDVAANVVEBPWKF
CKKLELKKNEKCVDGSTSQEVQWRRTVKKRDRHSKICTAQGPRDRMRSLQIARKFFD
LQDMLGFPDKASKTIEWLFSKSTSIKQLKERVAASEGGGKDEHLQVDEKEKDETLKLRVS
KRRTKTMESSEFKTKESRERARKRERERTMAKMKMRLFETSETISDPHQETREIKITNGVQ
LLEKENKEQEWSTNDVHMVEYQMDSVSIIKFLGLTSDSSSSSIFGDSSECYTSLSSVR
GTISAAGNSNVLTKNPN*
>G571 (326..1708)
TAGCCGACCTCTCTTCTCTTCTGAAAAAACACCAAAGGAGCTTTAAATGCTCCGTTA
CATAATCTCTATCTCTTTCCAAGAATATAGAGAAAGGAAAATAATATACAAGAATTAATA
GAAGGTATATCATCATCTCTCTAGCTAGTGATCAAAGCACCGTCATCATCATCATATATC
ATCAGCTTGCCCTCAGAGGAGAAGACCAACATAAGAGAGATCGAAGATCAAATCTATCTC
TCTTCATCATCTTCTGCTGTTACTATCATATCACACGCTCTCTCAAACATCATCTATAT
ATAGACTTCTCTTCATCATCATCAATGCAAGGTCATCACCAGAATCATCATCAACACTT
ATCATCATCCTCCGCCACGTCTTCCCATGGAACCTTCATGAACAAAGATGGGTATGATAT
TGGAGAGATAGACCCATCACTCTTCTCTATCTTGATGGACAAGGACATCATGATCCTCC
ATCAACTGCTCCTTCTCCTTTACATCATCATCACACAACTCAGAATTTGGCGATGAGACC
TCCAACATCGACGCTCAACATCTTCCATCTCAGCCTATGCACATAGAGCCACCTCCTTC
TTCTACACACAATACCGATAATACAAGATTAGTTCGGGCTGCTCAACCTAGTGGTTCCAC
TCGACCAGCTTCTGACCCGCTCCATGGACTTGACCAATCATTCTCAGTTTCATCAACCTCC
TCAAGGTTCTAAATCCATCAAGAAGGAAGGGAACCGCAAGGGTCTTGCTTCATCGGACCA
TGACATACCTAAATCGTCAGACCCTAAACATTGAGAAGACTAGCACAAAACAGAGAAGC
AGCAAGAAAAAGCAGATTACGTAAAAAGGCTTATGTTTCAAGCACTCGAGTCATGTAGGAT
CAAACCTGACCCAACTAGAACAAAGAGATTCAACGGGGCCAGATCCCAAGGCGTATTCTTTGG
AGGGTCTCTTATAGGAGGAGATCAACAGCAAGGTGGACTACCATTTGGCCCTGGCAACAT
CAGCTCTGAAGCAGCGGTGTTGATATGGAATATGCGAGGTGGCTGGAGGAGCAGCAGAG
GCTATTAAACGAACTAAGGGTGGCAACACAAGAACAACCTTGTCGAGAACGAGCTTAGGAT
GTTTGTGGACACATGTTTAGCTCATTATGACCATTGATTAACTCAAGGCTATGGTCGC
TAAGACCGATGTCTTCCACCTCATTTCTGGAGCATGGAACCTCCAGCTGAACGTTGCTT
CTTGTGGATGGGTGTTTTCCGTCCATCGGAGATCATTAAGGTGATTGTGAACCAGATAGA
ACCATTGACGGAGCAACAGATAGTTGGGATATGTGGGCTGCAACAGTCCACACAAGAGGC
CGAGGAGGCTCTCTCGCAAGGCTCGAGGCGTTGAATCAATCACTTCCGATAGCATTGT
CTCTGACTCCCTCCCGCTGCCTCCGCACCACTTCTCTCTCATCTATCCAATTTATGTC
ACACATGTCTTAGCTCTCAACAAGCTCTCTGCTCTCGAGGGCTTCGTTCTCCAGGCGGA
TAATTTGAGGCACCAAACGATCCATAGGCTGAACCAATTGTTGACGACCCGTCAAGAAGC
ACGGTGTCTTTAGCGGTTGCGGAGTACTTCCACCGTCTTCAAGCTCTAAGTTCTCTCTG
GCTAGCCCGTCTCGGCAAGATGGATAATACTAAAACAACTGATGAAGGAAACCAAAAAC
AAAAACAAGAGAATAGGTTGATTAGTTAGCCGCGAGCTTGACCTCTTTATCATATATATC
GTCTCTCTACTCAAATACAGTGCAATTAGGGAAAATTGTTTGGCTTCTTTTGGTATATG

ATTCTTACTATTATGTTTTTAATCAAGA

>G571 Amino Acid Sequence (domain in AA coordinates: 160-220)
MQGHHQNHQHQLSSSSATSSHGNFMNKDGYDIGEIDPSLFLYLDGQGHDPSTAPSPPLH
HHHTTQNLAMRPPTSTLNIFPSQPMHIEPPPSSTHNTDNLRLVPAAPQPSGSTRPASDPSM
DLTNHSQFHQPPQGSKSIKKEGNRKLGLASSDHDIPKSSDPKTLRRLAQNREAAKRSRLRK
KAYVQQLESCRIKLTQLEQEIQRARSQGVFFGSSLIGGDQQQGGPLPIGPGNISSEAAVFD
MEYARWLEEQQRLNLELRVATQEHLSENELRMFVDTCLAHYDHLINLKAMVAKTDVFLHI
SGAWKTPAERCFLWMGGFRPSEIIKVIVNQIEPLTEQQIVGICGLQQSTQEAEEALSQGL
EALNQSLSDSIVSDSLPPASAPLPPLSNFMSHMSLALNKLKSALEGFVLQADNLRHQTIH
RLNQLLTTRQEARCLLAVAEYFHLRLQALSSWLARPRQDG*

>G636 (6..1814)

CGATGATGCAACTGGGTGGTGGTACTCCGACCACTACAGCGGGGGCTACAACCGTCACAA
CTGCTACAGCACCACCGCCACAATCAAACAACAACGATTACAGCGCAACAGAAGCAGCGG
CAGCAGCGGTGGGGCGTTTGTAGGTGTGCGAAGAGATGCACGACCGTGGGTGGAGGAA
ATCGTTGGCCCGCGCAGGAAACGCTAGCGTTGTTGAAAATACGATCTGACATGGGAATAG
CGTTTCGAGACGCTAGCGTTAAAGGTCCCTTATGGGAAGAGGTTTCTAGGAAAATGGCGG
AGCATGGTTACATAAGAAACGCAAGAAATGCAAAGAGAAATTCGAGAACGTTTACAAAT
ACCAAAACGAACCAAGAAAGGTTCGTACCGGAAAATCCGAAGGCAAACTTATCGCTTCT
TTGATCAATTAGAAGCTCTCGAGTCTCAATCTACAACCTCACTCCACCATCATCAACAAC
AAACGCCTCTTCGACCACAGCAAAACAACAACAACAACAACAACAACAACAACAGCT
CCATATTTTCAACTCCTCCTCCGTTAAGCAGCTTATGCCGACGCTTCTCTTTCATCAA
TTCCTCCGTATACCTCAGCAGATTAATGTACCTTCGTTTCCAAACATCTCCGGTGATTTTC
TATCGGATAATTCTACATCGTCTTCGTCTTCTTATTCGACTTCTTCTGACATGGAGATGG
GTGGTGGAACCTGCGACTACAAGGAAGAAAGGAAGAAAGTGAAGAGTGTTCGAGC
GGTTGATGAACAAGTAGTTGATAAAGCAGGAAGAGCTTCAACGCACATTCTTGAAGCTG
TTGAAAAGCGAGAACACAAGAGATTGGTTAGAGAAGAGTCTTGAGAGATTCAAGAGATTG
CCAGAATCAACCGCAGCAGCAGATCTTAGCTCAAGAACGCTTATGTCCGCTGCAAAAG
ACGCTGCTGTTATGGCCTTTCTTCAAAAAGTGTGAGAGAAACAACCGAATCAGCCACAAC
CGCAGCCTCAGCCGCAACAAGTTCGACCATCAATGCAGCTTAATAACAACAATCAGCAGC
AACCCTCTCAACGGTCTCCTCCACCGCAACCTCCTGCTCCGCTTCCGAGCCAATTCAG
CGGTTGTGTCGACGTTAGACACAACGAAAACGCACAATCGTGGTGATCAGAATATGACTC
CTGCAGCTTCAGCGAGCTCGTCGCGGTGGCCGAAAGTGAGATAGAAGCATTGATAAAGC
TGAGGACGAATCTTGATTGCAAAATATCAAGAAAACGGACCAAAAGGACCATTGTGGGAAG
AGATATCAGCGGAATGAGAAGGTTAGGATTCAACAGGAACCTCAAAGAGATGCAAGAGA
AATGGGAAAACATAAACAATACTTCAAGAAAGTCAAAGAGAGCAACAAGAAACGTCCTCCG
AAGATTCCAAGACTTGCCTTACTTTCACAGCTTGATGCTTTATATAGAGAGAGGAACA
AATTCACAGCAACAACAACATTGCAGCTTCTTCTTCATCTTCCGGTCTTGTAAACCGG
ATAATTCTGTTCCCTTGATGGTCCAACAGAGCAGCAATGGCCTCCGGCTGTAACGACTG
CGACAACCTACTCCCGCAGCGGCTCAGCCTGATCAGCAATCTCAGCCGTCGGAGCAGAACT
TTGATGATGAAGAAGGTACAGATGAAGAGTACGACGATGAAGATGAGGAAGAGGAGAATG
AAGAAGAGGAAGGAGGTGAGTTCGAGCTTGTGCCTAGCAATAACAACAACAAGACGA
CGAATAATCTGTAATGATGATGATTGAGTTCGAACCGGTTTGGTGGTGAAAGATTAGTA
ATCTTTTTTTAAGTTTTGATACAGAACATGAGAATTTAAATATTGGAGGGTTT

>G636 Amino Acid Sequence (domain in AA coordinates: 55-145, 405-498)

MQLGGGTPTTTAAATTVTATAPPPQSNNNSAATEAAAAAVGAFAVSEEMHNRGFGGGR
WPRQETLALLKIRSDMGIAFRDASVKGPLWEEVSRKMAEHGYIRNAKKCKEKFENVYKYH
KRTKEGRTGKSEKTYRFFDQLEALESQSTSLHHHQQTPLRPQQNNNNNNNNNNSSSI
FSTPPPVTVMPTLPSSSIPPYTQQINVPSPFNISGDFLSDNSTSSSSSYSTSSDMMGG
GTATTRKKRKRKWKVFFERLMKQVVDKQELQRTFLEAVEKREHKRLVREESWRVQEIR
INREHEILAQERSMSAAKDAAVMAFLQKLSEKQPNQPQPQPQVPSMQLNNNNNQQP
PQRSPPPPAPLPPQPIQAVVSTLDTTKTHNRGDQNMTPAASASSSRWPKVEIEALIKLR
TNLDSKYQENGPKGPLWEEISAGMRRLGFNRSKRCKEKENINKYFKVKESNKKRPED
SKTCPYFHLDALYRERNKFHSNNNNIAASSSSGLVKPDNSVPLMVQPEQQWPPAVTTAT
TTPAAAQPDQSQPSEQNFDDEEGTDEEYDDEDEEEENBEEEGGEFELVPSNNNNNKTTN
NL*

>G878 (197..1738)

CAAAAAAATCTCTCCCATTAAGAACTGCCCAAAGAAATATTTTATACAAAATGAAAGA

GAGAAACACGACACGAATTTTGTATAATTAAGATTACACAAAAAAGTGTAGAAAGAG
AAATATCTTCTCTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTATCCTTTTGTTC
TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAAGAACCATCGAAGTTAAATCATC
CACCGGAGTTTACGGCCAACGATTTTACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
TAGCGGTGGCGTTGGATTTAGTCTTGGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
TCCTGATGAGTTCAAGTCTTCTCTCAGCTTTAGCTGGAGCTATGGCTTCTCCGGCGGC
AGCTGCTGTTGCCGCCGCTGCTGTGGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTTAAAGCAGAGTAGACC
AACCGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
GGCTACTCTTTTGGATTCTCCGAGCTTCTTGGTCTTTTTCACCTCTTCAGGGAACATT
TGGTATGACACATCAACAAGCTTTAGCACAAAGTCACTGCACAAGCAGTTCAAGGCAATAA
TGTTTCATATGCAGCAATCACAACAATCTGAATATCCTTCTTACACAACAACAACA
ACAACAACAACAGCTTCATTGACTGAGATTCCATCATTTTCTTCTGCACCTAGGTCTCA
GATTCGAGCCTCGGTTCAAGAAACATCCGAGGGTCAGAGAGAGACTTCGGAATATCTGT
CTTTGAGCATCGGTACAGCCTCAAAATGCTGACAAACCAGCTGATGATGGATACAACTG
GCGGAAATATGGGCAGAAGCAAGTGAAGGGGAGCGATTTCCTCGGAGTTATTACAAATG
TACGCATCCAGCTTGTCTGTCAAGAAGAAAGTGGAGAGGTCCTCGATGGACAAGTAAC
GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTCAAAAGCGCGGTAACAATAA
CGGGAGTTGTAAAGTTCTGATATTGCAAAATCAGTTTCAACAAGTAATAGCAGTCTCAA
CAAGAGTAAGAGGGACCGGAAACAAGCCAAAGTTACAACAACAGAGCAGATGTCTGAAGC
AAGTGATAGCGAGGAGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
GCCTGATCCCAAGCGAAGAAATACAGAAGTTCCGGGTTTCAAGAACAGTTGCTTCATCGCA
TAGAAGTGTGACAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
TGATGGATATAGGTGGCGCAAGTATGGTCAGAAAAGTAGTCAAGGAAATCCTTATCCGAG
GAGCTACTATAAGTGTACAACACCAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAAACATAACCATGATGTTCCAGC
TGCTAGAACCAGCAGCCATCAGTTAAGACCAAAACAATCAACACAACACCTCAACGGTTAA
CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGGCTTAAAGAAGAGCAAACTCACTTGACA
GAGAAGAAGAATACGACGGCGCTTGAGCTTTTGTGAGTTTAAATGAATCTTCTTTTGGTT
AATGAACCTGTTTTTGTGCTTCAAAAACACCACAGGTTTCTCTGGACAGAATCTCTGATA
TTACAGTTTCAAAAGGTATGTTCTTTTATTTCATGTTGGAATCTTCTGTGTAATCTTAAG
AAGCTTTAGGAGGTAAATGTAAAAAACCAGATTCAAAGTTATGCCCTTATGTGAATCTTT
TGACATGGGATAAACAATTTACAGGTATCCTTTTTTGTCTTGTGTAATAAAAAAAA
AAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKKEEPSKLSSTGVSRPTISLPPRPFGEFFSGGVGFSPGPMTLVSNLFSDPDEFK
SFSQLLAGAMASPAAAVAAAAVVATAHHQTPVSSVGDGGSGGDVDPFRFKQSRPTGLMI
TQPPGMFTVPPLSPATLLDSPSFFGLFSPLOQTFGMTHQQALAQVTAQAVQGNVHMQQ
SQQSEYPSSTQQQQQQQASLTEIPSFSSAPRSQIRASVQETSQGGQRETSEISVFEHRS
QPQNADKPADDGYNWRKYGQKQVKGSDFPKSYKCTHPACPVKKKVERSLDGQVTEIYK
GQHNHELPQKRGNNNGSKSSDIANQFQTSNSSLNKSIRDQETSQVTTTEQMSEASDSEE
VGNAEYSVGERHEDEPDPKRRNTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
RKYGQKVVKGNPYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
HQLRPNQHNSTVNFNHQPPVARLRLKEEQIT*

>G1134 (61..849)

TAAAGAAAGAGAAAAAAGCTTTTCGTAGTGTCTATTGAAACCAGAGAAAAGCCAAAGGGG
ATGCAACCAACATCCGTCGGTAGTAGCGCGGGTGGTGACGACGAGGAGGAGGAGGAGGA
GGAGGAGGGCTAAGTAGAAGTGGACTATCTCGGATCCGTTTCAGCTCCAGCGACTTGGCTT
GAAGCTTTACTTTGAGGAAGATGAAGAAGAGTCTTTGAAACCTAATCTTGGTCTCACCGAT
TTGCTTACCAGGAACTCGAACGATTTACCGACAAGTCGCGGCTCGTTTCAGATTCCCAGATT
CCTGTTGAGCAAGGGTTGATCAACAAGGTGGGTTTCACCGACAGAATAGTACTCCGGCG
GATTTTCTTGTGTTTCTGATGGATTATCCAAAGCTTTGGGATTGAGGCGAATTACGAT
TACTTATCGGGGAATATCGATGTTTCTCCGGGAAGTAAGCGGTCTAGAGAAATGGAAGCA
CTCTTCTCTCTCTGAGTTTACTTCTCAAATGAAAGGAGAGCAAAGCAGCGGTCAAGTT
CCTACCGAGTATCAAGCATGTCGATATGAACATGGAGAACCCTTATGGAGGACTCTGTT
GCTTTTAGGGTTCCGGCTAAACGTGGTTGCGCAACTCATCCCCGAGCATTGCCGAGAGG
GTACGAAGGACGCGGATTAGTGATCGGATAAGGAAGCTACAAGAGCTTGTACCTAACATG

GACAAGCAAACCAACTGCAGACATGTTAGAAGAAGCAGTAGAATACGTGAAAGTTCTT
CAAAGGCAGATCCAGGAGTTAACAGAAGAACAGAAGAGGTGCACATGCATACCTAAGGAA
GAACAATAAGGTTTGCTCCTGATTTGTTTATATTGCTTAACGGCAATGATCTGATCGA
AAAATTCGAAAGATGATCTTAGCTTGAATTTAGATGGATGTCATGTTGAAAAGTATATTA
TTTGATAAATGGATGTAGGTGTAATATAAAATTTTGTACAATAATGAAGAAAGTTAAAA
AGAATTAATGAAAACATATATTCTTTATGATATAAAAAAAAAAAAA
>G1134 Amino Acid Sequence (domain in AA coordinates: 198-247)
MQPTSVGSSGGGDDGGGRGGGGLSRSLRSRISAPATWLEALLEEDEEESLKP NLGLTD
LLTGNNSNDLPTSRGSFEFFIPVEQGLYQQGGFHRQNSTPADFLSGSDGFIQSFGIQANYD
YLSGNIDVSPGSKRSREMEALFSSPEFTSQMKGEQSSGQVPTGVSSMSDMNMENLMEDSV
AFRVRAKRGCAHPRSAERVRRTISDRIRKLQELVPMNDKQNTNTADMLEEA VEYVKVL
QRQIQELTEEQKRCTCIPKEEQ*
>G1008 (89..973)
GCCTTTTTGACTCTTCTTTCTCTCTTCTACTTTTTTTCAGGCTCTCTCTCTATATCTCTA
TCTTCTTCTCCGGTTAACTAAAAGAGAAATGAAAAGCCGAGTGAGAAAATCCAAGTACAC
GGTTCACCGGAAAATCACATCCACACCGTTCGACGGTTTCCCGAAGATTGTCAAATCAT
AGTCACTGACCCATGCGCTACTGATTTCTCCAGCGATGAGGAAAACGACAACAAATCTGT
TGCTCCGAGGGTGAAAACGTTATGTGGATGAGATCAGGTTCTGTGACGAAGATGACGAACC
TAAACCGCGAGGAAAGCGAAGAAAAGTCCCCGGCGGCTGCGGCGGAGAACGGTGGAGA
TTTGGTAAAGTCTGTGGTGAAGTATAGAGGAGTGAGACAACGACCTTGGGGAAAATTTGC
GGCGGAGATTCTGTGATCCTTCGAGTCGTACTAGACTCTGGCTTGGGACTTTTGCGACGGC
GGAGGAAGCTGCTATAGGTTACGATAGAGCCGCGATTTCGAATCAAAGGTCATAACGCTCA
GACGAATTTTCTACTCTCTCTCTAGTCCGACGACTGAGGTGTTACCGGAACTCCGGT
GATTGACCTTGAAACTGTCTCTGGTTGTGATTCGGCGAGGGAATCGCAAATCAGTCTGTG
TTCTCCGACTTCTGTTCTCCGGTTTAGTCACAACGACGAAACAGAGTACAGAACAGAGCC
AACGGAAGAACAAAATCCGTTTTTCTTGCCTGATTTGTTTCGCTCCGAGATTATTTTTG
GGATTCCGAAATTACCCCTGACCTTTGTTTCTCGACGAATCCACCAGTCTTGTACC
AAACATCAACAACAACAACACAGTGTGTGATAAGGATACGAATCTGTCTGATAGTTTTCC
GTTGGGAGTGATCGGAGATTTAGCTCATGGGATGTTGATGAGTTTTTCCAAGATCATT
GTTGGATAAGTAATTTGATGAGTCTTCCCCAGAATTTTCTGGGTTTCTCTTTTGGTT
GTGTGAGTGAATGAGTGGTTTGATGACAACGACGGGGATGAATCTTAGCCGTCGGTTTT
CCATTTCTGTTGGACGGCTCCGATCAGCGGAAGAAGCGCAACGGAGTTTTTATTTATCTGTT
TGAGAATTTTATAATTTAATTTGCGAGTAAATATAGTAATTAGTGTAAAGATTGTGAGAG
TTTAAGTTAATTAGGGAGGGGTTTTGAATATTGGGGATTTTGGGAGGTTTTGTTTGGTT
TCTCTCCAAGTCTGTCACTATGCAAGGAAGCAGTATAAAGACCGTATATATATTTATTA
TTAATATTGATAAAAGTAAAAAAAAAAAAAAAAAAAA
>G1008 Amino Acid Sequence (domain in AA coordinates: 96-163)
MKSRVRKSKYTVHRKITSTPFDGPKIVKIIVTDPCATDSSSDEENDNKSVA PRVKRYVD
EIRFDEDEDEPKPARKAKKKS PAAAAENGDLVKS VVKYRGVRQRPWGKFAAEIRDPSSR
TRLWLGT FATAEEAAIGYDRAAIRIKGHNAQTNFLT PPPSPTEVLPETPVIDLETVSGC
DSARES QISLCSPTSVLRF SHNDETEYRTEPT EEQNPFFLPDLFRSGDYFWDSEITPDPL
FLDEFHQSLLPNINNNTVCDKDTNLSDSFPLGVIGDFSSWDVDEFFQDHL LDK*
>G1020 (132..689)
CTGTTTCAAGAAAGCTCCCCAAAAGGAGCGTTGCTTTACTCTCTCTATAAAAAGAGCTC
TTCTACTTCTTCTCTGTTTACCACAAAACCTTTTACCAGATCTTCTCGTTCCATTCTTCTC
CTAATTACACCATGCCCCAACATCACCATGGGTTTGAAACCCGACCCGGTTGCTCCAACGA
ACCCGACTCATCATGAGAGTAATGCTGCCAAAGAGATTCGTTACAGAGGCGTTAGGAAAC
GTCCATGGGGAAGATACGCCGCTGAGATCCGAGATCCGTTAAGAAAACCTCGAGTCTGGC
TCGGTACGTTTCGACACCGCTCAGCAGGCGGCGCGTGCTTACGACGCGAGCCGCGCGTACT
TTCGTGGTGTTAAGGCTAAGACCAATTTCCGTGTTATCGTTGGTAGTAGTCTCTACTCAGA
GTAGCACCGTCTGCTGACTCTCCACGGCGGCGCGGTTTATAACACCTCCGCACCTCGAGC
TCAGCTTAGGCGGCGGCGGCGCGGTGTCGTCTGTAAGATCCCGCTTGTCATCCGTTTACT
ACTATAACATGGCGACGTATCCAAGATGACGACGTGTGGTGTCCAGAGCGAGTCTGAAA
CGTCTGTCGGTCTGTTGATTTTGAAGGTGGAGCTGGGAAGATATCTCCGCCGTTAGATCTGG
ATCTTAACTTAGCTCTCCGGCGGAATAGGCCGTGAGTTTTTTTTTCTTATGTCGTTTC
TTTAGACAAAAAAATAACGTTTCTTTTTTTTCTGCTAAGAAAAAATATTTATCCG
TTTTTTAGAAGAAAAAATAAAAAAAAAAAAAAAAAAAAA

>G1020 Amino Acid Sequence (domain in AA coordinates:28-95)
MPNITMGLKPDVPVAPNTNPHHESNAAKEIRYRGVVRKRPWGRYAAEIRDVPKKTRVWLGT
DTAQQAARAYDAAARDFRGVKAKTNFGVIVGSSPTQSSTVVDSPTAARFITPPHLELSLG
GGGACRRKIPLVHPVYYNMATYPKMTTCGVQSESETSSVDFEGGAGKISPPLDLNL
APPAE*

>G1023 (252..1250)

TCGTCTTCTTAATCGCTTTCTGCTCTGTTTTCTCGTTCATCAAGCTACATCTACTAGCT
CTCTCAGTGATGTATTTCTCACAGTTTCATCGATTTCATGCGTTTAAAGACCTAAAAGGA
CTTGTCTGSGGTAAAGGACTTTTCTGTCTTGAGAGAGTTCATTTTGAGGCTTTTCTG
GGAATTTTGAGAGGTTTTTAGGGTTTAAAGGGGTTTGGTTTTGAATTCGCACACCAAG
TGTTTCGATAAATGGCTGAACGAAAGAAACGCTCTTCTATTCAAACCAATAAACCAACA
AAAAACCCATGAAGAAGAAACCTTTTTCAGCTAAATCACCTCCCAGGTTTATCTGAAGATT
TGAAGACTATGAGAAAACCTCGTTTCGTTGTGAATGATCCTTACGCTACTGACTACTCAT
CAAGCGAAGAAGAAGAAAGGAGTCAGAGAAGGAAACGTTATGTCTGTGAGATCGATCTTC
CTTTCGCTCAAGCTGCTACTCAAGCAGAATCTGAAAGCTCATATTGTGAGAGAGTAACA
ATAATGGTGTAGCAAGACTAAAATCTCAGCTGTAGCAAAAAGGTTTACGCAGCAAAG
CATCTCCGGTCGTTGGACGTTCTTCTACTACTGTCTCGAAGCCTGTTGGTGTAGGCAGA
GGAAATGGGGTAAATGGGGTGTGAGATTAGACATCCATCACCAAAGTAAGAACTTGGT
TGGGTACTTACGAGACGCTGAACAAGCAGCTGATGCTTATGCTACCAAGAAGCTTGAGT
TTGATGCTCTGGCTGCGACCACTTCTGCTGCTTCTCTGTTTTGTCAAATGAGTCTGGTT
CTATGATCTCAGCCTCAGGGTCAAGCATTGATCTTGACAAGAAGCTAGTTGATTGCGACTC
TTGATCAACAAGCTGGTGAATCGAAGAAAGCGAGTTTTGATTTGACTTTGCAGATCTAC
AGATTCTGAAATGGGTGCTTTCATTGATGACTCATTTCATCCCAATGCTTGTGAGCTTG
ATTTTCTCTTAACAGAAGAGAACAACAACCAATGTTGGATGATTACTGTGGCATAGATG
ATCTGGACATCATTGGTCTTGAATGTGACGGTCCAAGCGAATTCAGACTATGATTTCT
CAGATGTGAGATCGATCTTGGTCTCATTTGAACACCATTGACAAGTATGCTTTCGTTG
ATCATATCGCAACAACACTACTCCACTCCTCTTAATATCGCGTGCCCATAGTTTTCGAGC
TAGGTGTTATTATTAGCTATAGGAGCAACGTAAGGCTCGTTGTTACTCGGTTTTGTCT
TAAGTTATTAAAGTATAGCAGAGGCAGTTAATCTCAAGGGAAGCAAAAACCTAAAGATA
GAAGCAGATGCAGTTTTGTGTGTTGGTGTACTAAAGAAAGTTTTGTGACATAATGGTT
TTGATGTTGTGAGAGAAGATAGAGAGGTGTGATCGAAATTGTAATCTCAGGTGGTTTTTT
TTGAAGGCAATTGTTCTCATTAGGGTTTTTTTCTATATGAGGATTGTCTTTGAAAAGC
CTTTAGATGTTTTCTAATTCGTAAGCTCTCTCAATCTTTGTAAGTTTTGCCTGTTGAGTT
ATTGATACATATGTGAGACCTACTTTATTTGTTTTGTGCTACATACATTGTTGATGGTT
CGTCAAAAAAAA

>G1023 Amino Acid Sequence (conserved domain in AA coordinates:128-195)
MAERKRSSIQTNPKNKPKMKKPFQLNHLPLGLSEDLKTMRLRFVNDPYATDYSSSEE
EERSQRRKRYVCEIDLFFAQAATQAESESSYQESNNNGVSKTKISACSKVLRSKASPV
VGRSSTTVSKPVGVRQRKWGWAAEIRHPITKVRTWLGTYETLEQAADAYATKKLEFDAL
AAATSAASSVLSNESGSMISAGSSIDLKKLVDSLTLDQAGESKKASFDFDLQIPE
MGCFFIDDSFIPNACELDFLLTEENNNQMLDDYCGIDDLDIIGLECDGPSELDPDYDFSDVE
IDLGLIGTTIDKYAFVDHIATTTPTPLNIACP*

>G1053 (38..538)

GAAACTCTTACATACTCATATAAACCAAACTAAAACCATGATTCCGGCAGAAATCAACGG
ATATTCCAATATCTATCACCGGAATACAACGTAATAAACATGCCTTCATCTCCAACTC
TTCTTAAACTACCTAAACGATTTGATCATCAACAACAACACTATTCTCATCATCCAA
CAGTCAAGATCTCATGATAAGCAACAACCTCAACTCCGACGAAGATCATCATCAAGCAT
CATGGTACTCGACGAGAGGAAACAGAGAAGGATGCTTTCGAACAGAGAATCTGCAAGGAG
GTCAAGGATGAGGAAACAGAGACATCTTGATGAACTCTGGTCTCAGGTAATAAGGCTTCG
CAACGAGAACAACCTGTCTTATCGATAAGCTGAACCGGTATCGGAGACTCAAAATTGTGT
ATTGAAGGAGAACCTTAACTCAAAGAAGAAGCTTCTGATCTCCGACAGCTTGTGTTGTA
ACTGAAATCTAACAAGAACAACAACATAGTTTTCCAAGAGAGTTTGAAGATAATTAGTA
TTACTCAA

>G1053 Amino Acid Sequence (domain in AA coordinates: 74-120)
MIPAEINGYFQYLSPEYNVINMPSSPTSSLNYLNDLIINNYYSSSSNSQDLMISNNSTS
DEDHQSIMVLDERKQRRMLSNRESARRSRMRKQRHLDELWSQVIRLRNENCLIDKLR
VSETQNCVLKENSCLKKEASDLRQLVCELKSNKNNNNSFPREFEDN*

>G1137 (202..1248)

TACTTCAGACTTCTACTCAAACCAAGTCACGTAGTTGGTTGGTGACATTTTCGCTGCATTTT
TCAATCTGTGATTGTTTTTCGTTTCGTCTTTCTTTTACTATTTCCTCGAAAAGGACACAAG
AAGTATTGCATTCACTCAGTTGAGCAACTTAACAATCGTGTGTACTTTTTGAAGTTCCC
TTGAGCTAAACTGCTAAGAGCATGCCTCTGGATAAGAGGCAACGGGATTTGCCTCTGGGC
TTAAGTCCTCAAGCTTGCTTCAAGGATATAGTAGGTCGGTCTGTCTTCCTAGAATTCCT
CTCCCTGAGCTTGGGAACTATATGCAGCTAAGCTTCAGGCTCGCTGTTTGCAGCCACCA
CCATTCCAGTCTTTGCTGTGCAGTCATGATAAGGAGTCTTATGGAAAAAGATTCTCACGG
TCTGACATGCGGTCTTGGTGCGCTGCTGCTACTACTACTACTACTCCACTTGGAGCATT
GAGTCTTCTCAGAAAAGACTTTTGATATTTCGATCAGTCAGGAGACCAGACTCGTCTATTA
CAATGTCCATTTCTCTACGGTTTCCATCTCATGCGGCTGCAGAACCAGTGAAACTCTCT
GAGTTACAAGGTATAGAGAAAGCTTTTCAAAGAAGATGGTGAAGAGTTTACAAGAGTGAT
GGAACAGAGTCAGAAATGCATGAAGACACTGAGGAGATCAATGCATTGCTATATTCAGAT
GATGATTATGATGATGATTGCGAGAGTGATGATGAAGTAATGAGCACTGGTCACTCTCCT
TATCCAAATGAAGGAGTTTGCAACAAAAGGGAATTAGAAGAAATCGATGGTCTTGTA
AGGCAGAACTACTGGATAAGGTCAACAACATCAGCGACTTATCATCACTTGTGGGCACT
GAGAGCTCCACACAACCTCAATGGATCTTCTTTCTTAAGGACAAAAGCTCCCTGAATCA
AAAACCATATCGACCAAAGAGGACACTGGTTCTGGTCTGAGCAACGAGCAGTCGAAGAAA
GACAAGATCCGACAGCTCTGAAAATACTCGAGAGCGTAGTCCCTGGTGAAAAGGAAAC
GAAGCGCTCTTACTTCTGGACGAAGCAATTGATTACCTAAAGTTGCTGAAACGAGACTTA
ATCTCCACAGAGGTTAAGAACCAAGCTCCACCACTACAAGTCACCAATCTTGTGCTT
AAAGAGACAACATGGGGACAAGAAATCTGCAGACAGATAAGGCGTGAAAGATTCTGACG
AGTTAAAACCTGTGAAGTGGGTTTTTGGGTACGTATCCTTGCACCAGCTTT

>G1137 Amino Acid Sequence (domain in AA coordinates: 264-314)

MPLDKRQRLDPLGLSPQACFKDIVGRSVLPRIPLPELGKLYAAKLQARCLQPPPFQSLLC
SHDKESYGRFSRSDMRWCAAAATTTTPLGALESSQKRLILFDQSGDQTRLLQCPFPLR
FP SHAAAEPVKLSLEQIEKAFKEDGEEFHKSDGTESEMHEDTEEINALLYSDDDYDDDC
ESDDEVMSGTGHPYPNEGVCNKRELEIDGPCKRQKLLDKVNNISDLSSLVGTESSTQLN
GSSFLKDKKLPESKTISTKEDTGSGLSNEQSKDKIRITALKILESVPVPAKAGNEALLLLD
EAIDYLLKLLKRLISTEVKNQSSSTTHKSPILLKETTGWTRNLQTDKA*

>G1181 (113..1012)

CTCGATCTTTTAACCCCATTTATTACATATTACTCCTTCCTACATTATTCTTCTTCTGCT
TTCGTGACTTTTCAGGGGACACTTTTGTTTTATAACTTACGCTTAAAACTCCTATGAATTC
GCCGCCGGTTGACGCAATGATTACCGGAGAATCATCGTCACAAAGATCTATCCCAACGCC
GTTTCTCACAAAACGTTTAACTCGTTGAAGATAGTTCATCGACGATGTTATCTCATG
GAACGAAGATGGTTCCTCTTTTCATCGTATGGAATCCGACAGATTTTCGCTAAAGATTTGCT
TCTTAAACACTTCAAACACAACAATTTCTCTAGTTTCGTTTCGTCAGCTCAACACTTACGG
ATTCAAAAAGTTGTACCGGATCGATGGGAGTTTCAAACGATTTCTTTAAGAGAGGAGA
AAAACGTCTTCTCCGTGAGATCCAACGTCGAAAATAACAACGACGCATCAAACAGTTGT
TGCTCCTTTCGTCGGAACAACGAAACAGACGATGGTTGTATCACCGTCAAATTCGGGGA
AGATAATAATAATAATCAGGTGATGTCTTCGTCTCCGTCTCGTGGTATTGTCATCAAAC
GAAGACGACTGGGAATGGTGGTTTATCAGTGGAGTTATTGGAAGAGAACGAGAAGCTTCG
GAGTCAAAACATTAGCTAAACCGTGAGCTTACTCAGATGAAATCTATCTGCGATAATAT
CTATAGTCTCATGTCCAATTACGTCGGATCTCAGCCCACTGATCGGAGTTATTCTCCCGG
AGGTAGTAGTAGTCAACCGATGGAGTTTTTACCGGCGAAGCGGTTTTTCGGAGATGGAGAT
TGAAGAAGAAGAAGAAGCGAGTCCGAGGTTGTTGGTGTTCGATTGGGTTAAAACGGAC
GAGAAGTGAAGGTGTTCAAGTGAAGACGACGCGGTTGGTGGGAAAATTCCGATGAGGA
GACGCCGTGGTTGAGACATTATAATCGAACCAATCAGAGAGTTGTAATTAACGAAC
GGTTTAGATTTGTGGTGTAGATATGTGCGCGAAGTAGACGATTACAGCTTTTAAAGACAA
GCAGAGCACGTGTCCATCTGTTTCAAGAAGTTCTGCAATCTTGACTTCTTCTTTTAAAC
ACTTTGTGTTTTTTTATTATTTAATAAATAAATAAATGTTCTTTTTCAGTTTGTGTTTT
TTCAAAAATAGTTCGGCTGTTTCTAGACTTTCCTTTTTT

>G1181 Amino Acid Sequence (domain in AA coordinates: 24-114)

MNSPPVDAMITGESSQRSIPTPLTKTFNLVEDSSIDDVISWNEDGSSFIWNPDTFAK
DLLPKHFKHNNFSSFVRQLNTYGFKKVVPDRWEFSNDFFRGEKRLLEIQRKITTTHQ
TVVAPSSSEQRNQTMVSPNSGDNNNNQVMSSSPSSWYCHQTKTNGGLSVELLEENE
KLRSQNIQLNRELTMQMSICDNIYSLMSNYVGSQPTDRSYSPGSSSQPMEFLPAKRFSE

MEIEEEEEASPRFLGVPVIGLKRTRSEGVQVKTAVVGENSDEETPWL RHYNRTNQRVCN*
>G1228 (63..1139)
GCATTTATAATTACTCACTCATCTTCTTTTCATTACATTACATACCAAACAAGAGCTCTC
AAATGGAAGGTTTCAAGGACACATCAACCCCTGTTCTTCGATCGAAAACCGGATGTGA
GAAGCCTCGAGGTTCAAGGATTTGCAGAGGCTCAAAGCTTGTCTTCAAAGAAAAGAGG
AAGAAAGCTTACAAGATACAGTTCCATTCTACAGATGCTGCAAAGTGAAGACCCCTCAT
CGTTTTTTTCAATCAAAGAGCCAACTTTCTGACGCTACTGTCTCTTCAAACCCCTCAAGG
AGCCTTGGGAACTCGAAAGATATCTTCACTTGAGGATTACAAATTTCACTTACCAGGTCC
AATCTGAGACCAACCGCTTCATGGAAGGAGCCAATCAAGCTGTGTCAAGCCAAGAAATTC
CCTTTAGCCAAGCAAACATGACACTCCCTTCTTCTACCTCATCACCCTCAGTGCACATT
CAAGACGAAAGCGCAAAATCAACCCTTGCTGCCTCAAGAAATGACTAGAGAAAAGAGAA
AGAGGAGGAAAACAAAACCAAGTAAAAACAATGAAGAGATTGAGAATCAAAGAATAAACC
ACATTGCTGTTGAACGAAACAGAAGACGTCAAATGAACGAACATATCAACTCTCTCCGGG
CCCTTCTCCCACTTCTTACATCCAACGAGGAGACCAAGCTTCCATAGTAGGAGGAGCAA
TAACTACGTGAAGGTCCTCGAGCAAATCATACAATCTCTCGAATCGCAAAAGAGAACGC
AACAAACAAAGTAACAGTGAGGTAGTAGAAAACGCACCTTAATCATCTCTCAGGCATTTCTG
CGAACGACCTGTGGCAACTCTTGAAGATCAAACCTGTATCCCCAAAATCGAAGCTACAG
TGATACAAAACCATGTGACGCTTAAAGTTCAATGTGAGAGAAACAAGGACAACCTTCTCA
AAGGAATCATATCACTTGAAAAGCTTAAACTCACTGTTCTTCATCTCAATATCACTACTT
CGTCTCATCTCTCTGTTTCTTATCTTCAACCTCAAGATGGAAGATGAGTGCACCTTAG
AGTCAGCCGACGAGATTACGGCGCTGTTTCATCGGATTTTCGATATTCGACAATTTGAT
TAAACACATATAATTCCAAAATATTAACAGCTGACAAAATGGTATCTTTGCGGCC
>G1228 Amino Acid Sequence (domain in AA coordinates: 179-233)
MERFQGHINPCFFDRKPDVRSLEVQGFQFAEQSFAPKEKEEESLQDTPVPLQLMQSEDPSS
FFSIKEPNFLTLTSLQTLKEPWELERYLSLEDSQFHSVPQSETNRFMEGANQAVSSQEI
FSQANMTLPSSSTSSPLSAHSRRKRKINHLPPQEMTREKRKRKTKPSKNNEEIEINQRINH
IAVERNRRRQMNHINSLRALLPPSYIQRGDQASIVGGAINYVKVLEQIIQSLESQKRTQ
QQNSEVVENALNHLSGISSNDLWTTLEDQTCIPKIEATVIQNHVSLKVQCEKKQGQLLK
GIISLEKLKLTVLHLNITSSSHSSVSYSFNLKMEDECDLESADETAAVHRIFDIPTI*
>G1277 (51..512)
ATTCTAAAGTCCTCTCTCGGAAAGTAAGAGACTCAACTTCCGAGCCGCCATGGACGCCG
GAGTAGCAGTAAAAGCTGACGTGGCAGTCAAATGAAGAGAGAAAGACCATTCAAAGGGA
TCAGAATGAGAAAATGGGGGAATGGGTTGCGGAGATTCGAGAACCCAACAAGCGTTCAA
GACTTTGGCTCGGCTCTTACTCTACTCCGAAGCGCGCGCGTGCATACGACACGGCTG
TCTTTTACCTCAGAGGACCAACTGCTACGCTCAACTTCCCGAGCTTCTGCGGTGTACCT
CCGCCGAGGATATGTGACGCGCAACGATCAGGAAAAAGGCGACGAGGTGGGAGCTCAAG
TAGATGCGATAGGGGCGACGGTGGTGCAGAACAAACACCGCCCGCGGTTTTTAGTCAA
AGCGTGACTTTGGCGCGGGTTATTAGAGCTTGTGACTTGAACAAGTTACCTGACCCGG
AAAATCTCGATGATGATTGGTGGGAAAATAGACTGAAAAATAATAATAAATATCTTAC
AATGGTGGCTGTAGCTATCGTACGCGGAATGCTTGGGCTTGTGTATATGACTACGTGGT
TACGGAAGATTCTCTGTTTCGTATGTATTAAAAATTAATCCACAAGTCAAACATA
CTGTACATTATTCTTAATTTAGTATTTTCTTATTAATATCTATCATTTGTTTGGTGAACA
CCAGAATATTAGACTATTAAATGTAACGAGTTTTTAATATTTTCGATCATAATAACACCAAG
CTAGTTAAAGGTTAATATCTTGTGTTACGAGTCTTGAGTAAGTTCAATTGTCATATATG
TAACGGAAGAGGTTCTGTTCCGGTCCCAAGTGAAGTGATCAAAGGTGACTTCACATAAAA
AATAAAAAAAA
>G1277 Amino Acid Sequence (domain in AA coordinates: 18-85)
MDAGVAVKADVAVKMKRERPFKGIRMRKWGWAEIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTATLNFPELLPCTSAEDMSAATIRKKATEVGAQVDAIGATVVQNNKRRRV
FSQKRDFGGGLLELVBLNKLDPENLDDDLVGK*
>G1309 (53..859)
CGTCGACCTCTTAATTAAGACGACTTGAGAGAGAAAGAAAGATACGTGGAAGATGACCAA
ATCTGGAGAGAGACCAAAAACAGAGACAGAGAAAGGGTTATGGTCACCTGAAGAAGACCA
GAAGCTCAAGAGTTTCATCTCTCTCGTGGCCATGCTTGTGAGGACCTGTTCCCATCT
AGCTGGATTGCAAAGGAATGGGAAAAGCTGCAGATTAAGGTGGATTAAATTACCTAAGACC
AGGACTAAAGAGGGGGTCTGTTTAGTGAAGAAGAAGAAGAGACCATCTTGACTTTACATTC
TTCCTTGGGTAACAAGTGGTCTCGGATTGCAAAATATTTACCGGAAGAACAGACAACGA

GATTAAGAACTATTGGCATTCTCTATCTGAAGAAGAGATGGCTCAAATCTCAACCACAACT
CAAAAGCCAAATATCAGACCTCACAGAATCTCCTTCTTCACTACTTCTTGCGGGAAAAG
AAATCTGGAAACCGAAACCTTAGATCACGTGATCTCCTTCCAGAAATTTTCAGAGAATCC
AACTTCATCACCATCCAAAGAAAGCAACAACAATGATCATGAACAACAGTAATAACTT
GCCTAAACTGTTCTTCTCTGAGTGGATCAGTTCTTCAAATCCACACATCGATTACTCCTC
TGCTTTTACAGATTCCAAGCACATTAATGAAACTCAAGATCAAATCAATGAAGAGGAAGT
GATGATGATCAATAACAACAACACTACTCTTCACTTGAGGATGTCATGCTCCGTACAGATTT
TTTGAGCCTGATCATGAATATGCAAATTTATTATTCTTCTGGAGATTTCTTCATCAACAG
TGACCAAAATATGCTCTAAGAAGAGTGAATATGATCGTAAGAGGAACATAAGCTAGTTAC
TTGTGTTACAGC

>G1309 Amino Acid Sequence (domain in AA coordinates: 9-114)
MTKSGERPQKQKGLWSPEEDQKLKSFILSRGHACWTTVPILAGLQRNGKSCRLRWINY
LRPGLKRGSPFSEEEETILTLHSSLGNKWSRIAKYLPGRDNEIKNYWHSYLKKRWLKSQ
PQLKSQISDLTESPSSLLSCGKRNLLETLDHVISFQKFSENPTSSPSKESNNNMIMNNS
NNLPKLFSEWISSSNPHIDYSSAFTDSKHINETQDQINEEVMMINNNYSSLEDVMLR
TDFLQPDHEYANYSSGDFFINSDQNYV*

>G1314 (1..990)
ATGGGAAGAGCTCCGTGTTGCGACAAGACAAAAGTGAAGCGAGGGCCTTGCTCGCCTGAA
GAAGACTCTAAACTTAGAGATTACATTGAAAAGTATGGTAATGGTGGAATTTGGATCTCT
TTCCCCCTCAAAGCCGGTTTGAGGAGATGTGGGAAGAGTTGTAGACTGAGGTGGCTAAAC
TATTTGAGACCAAACATAAAGCATGGTGACTTCTCTGAGGAAGAAGACAGGATCATTTTT
AGTCTCTTCGCTGCCATAGGAAGCAGGTGGTCAATAATAGCAGCTCATCTACCGGGACGA
ACAGACAACGACATAAAAACTATTGGAACACAAAGCTAAGGAAGAACTCTTGCTCTCT
TCCTCTGATTCATCATCATCAGCCATGGCTTCTCTTATCTAAACCCTATTTCTCAGGAT
GTGAAAAGACCAACCTCACCAACAACAATCCCATCTTCTTCTTACAATCCGTATGCTGAA
AACCTTAATCAATACCCAACAAAATCCCTCATCTCCAGCATCAATGGCTTCGAAGCTGGT
GACAAAACAGATAATTTCTATATTAACCCTAATTATCCTCAAGATCTCTATCTCTCGGAC
AGCAACAACAACCTCGAAGCAGCAATGGTTTCTTGCTCAACCACAATATGTGTGATCAG
TACAAGAACCACACAGTTTTTCTTCAGACGTCATGGGATAAGATCAGAGATTATGATG
AAGCAAGAAGAGATAATGATGATGATGATGATAGACCACCACATTGACCAGAGGACAAAA
GGGTACAATGGGGAATTCACACAAGGGTATTATAATTACTACAATGGGCATGGGGATTG
AAGCAAATGATTAGTTAGTGGAAACAGCACTAATTCTAACATAAACATGGGTGGTTGAGTTCA
TCTTCTAGTTTCGATAAGCAACCTAGCTGAGAACAAAAGCAGTGGTAGCCTCCTACTAGAA
TACAAATGCTTGCCCTATTTCTACTCCTAG

>G1314 Amino Acid Sequence (domain in AA coordinates: 14-116)
MGRAPCDKTKVKGFPWSPEEDSKLRDYIEKYNGGNWISFPLKAGLRRCGKSCRLRWLN
YLRPNIKHGFSEEDRIIFSLFAAIGSRWSIIAHLPGRTDNDIKNYWNTKLKRLKLLSS
SSDSSSSAMASPYLNPISQDVKRPTSPPTIPSSSNPYAENPNQYPTKSLISSINGFEAG
DKQIISYINPNYPQDLYLSDSNNTSNANGFLLNHNMDQYKNHTSFSSDVNGIRSEIMM
KQEEIMMMMHIDHIDQRTKGYNGEFTQGYNYNNGHDLKQMSGTGTNSNINMGSGS
SSSSISNLAENKSSGSLLEKCLPYFYS*

>G1317 (1..849)
ATGGGAAGATCACCTGTTGTGATAAAAAATGGAGTGAAGAAGGGACCATGGACTGCTGAG
GAGGATCAGAAACTCATCGATTATATTCGATTTTCATGGTCTCGGCAATTGGCGTACGCTC
CCCAAAAATGCTGGACTCCATAGATGTGGAAGAGCTGCCGTCTTCGATGGACCAATTAT
CTAAGACCGGACATCAAGAGAGGAAGATTCTCGTTTCGAGGAAGAAGAACTATCATTAG
CTACACAGTGTTATGGGAAACAAGTGGTCAGCAATAGCCGCTCGTCTACCAGGGAGGACC
GATAACGAAATAAAAAACCATTTGGAACACTCACATCCGCAAGAGACTTGTAAGGAGTGGT
ATCGACCTGTTACTCATTCTCCACGCCTTGATCTTCTTGATTGTCTCACTTTTGAGT
GCACTTTTCAACCAGCCAACTTTTCAGCAGTTGCAACACATGCGTCTTCTCTCTTAAT
CCTGATGTATTGAGGTTGGCCTCTCTACTACTGCCACTTCAAAACCTAATCCAGTTTAC
CCATCGAACCTCGACCAAAATCTTCAAACTCCAAATACATCATCAGAATCGTCTCAACCA
CAAGCTGAGACTAGTACAGTCCCAACAACTATGAAACTTTCATCATTGGAGCCTATGAAC
GCAAGACTCGACGAGTGGTCTTGCGAGATGTATTACCACCTTTGTGAGAGATTTTGAC
TTAGACTCGCTCATGTCAACGCCAATGTCTTCTCCACGACAAAATAGCATTGAAGCAGAA
ACCAACTCCAGCACTTTCTTCGACTTTGGAATTCCGGAAGATTTTCATCTTAGATGACTTT
ATGTTTTTAA

>G1317 Amino Acid Sequence (conserved domain in AA coordinates:13-118)
MGRSPCCDKNGVKGPWTAEDQKLIDYIRFHGPGNWRTLPKNAGLHRCGKSCRLRWNTNY
LRPDIKGRGFSFEEETIIQLHSVMGNKWSAIAARLPGRTDNEIKNHNTHIRKRLVRS
IDPVTHSPRLDLLDLSSLLSALFNQPNFSAVATHASSLLNPDVLRASLLPLQNPVPV
PSNLDQNLQTPNTSSSQPQAETSTVPTNYETSSLEPMNARLDDVGLADVLPLPSESFD
LDSLMSTPMSSPRQNSIEAETNSSTFFDFGIPEDFILDDFMF*

>G1323 (49..870)
AAGAGGGAATCTCAAAAGTGTGTCTGTGTGAGAGAGGAGAGAGAATATGGGCAAAGGA
AGAGCACCATGTTGTGACAAAACCAAGTGAAGAGAGGACCATGGAGCCATGATGAAGAC
TTGAAACTCATCTCTTTTCATTACAGAATGGTCATGAGAATTGGAGATCTCTCCCAAAG
CAAGCTGGATTGTTGAGGTGTGGCAAGAGTTGTCGTCTGCGATGGATTAATTACCTCAGA
CCTGATGTGAAACGTGGCAATTTCAAGTGCAGAGGAAGAAGACACCATCATCAAACTTCAC
CAGAGCTTTGGTAAACAAGTGGTCAAGATTGCTTCTAAGCTGCCTGGAAGAAGACAGACAAT
GAGATCAAGAATGTGTGGCATAACATCTCAAGAAAAGATTGAGCTCGGAACTAACCTT
AATGCCGATGAAGCGGGTTCAAAGGTTCTTTGAATGAAGAAGAGAAGTCTCAAGAGTCA
TCTCCAAATGCTTCAATGTCTTTTGTCTGGTTCCAACATTTCAAGCAAAGACGATGATGCA
CAGATAAGTCAAATGTTTGAAGCATTCTTAAGTATAGCGAGTTTACGGGGATGTTACAA
GAGGTAGACAAACAGAGCTGCTGGAGATGCCTTTTGAATTAGATCTGACATTTGGAGT
TTCATAGATGGTTCAAGATCATTTCAACAACAGAGAACAGAGCTCTTCAAGAGTCTGAA
GAAGATGAAGTTGATAAATGGTTTAAGCACCTGGAAAGCGAAGTCTGGGTTAGAAGAAAAC
GATAACCAACAACAACAACAGCATAAACAGGGAACAGAAAGTGAACATTCATCATCACTC
TTGGAGAGTTACGAGCTCCTCATACATTAATGAAGCCATAAAGCAAGTCAATTTTACCTT
GAAAATGGAATTATTAGCTAAGTATTGTCATTATTAGTATATAAGCAAGATCAGATAGG
CGCATGTAGTAGCAACAAGAAACGTCGAATTGTAGACAAAATGTAGATATTACAGA
GTTGAAAAGATTGATTTTGCAAATGATTGCTTTGTAGTGAAATCAAGTTATCACAAAAA
AAAAAAA

>G1323 Amino Acid Sequence (domain in AA coordinates: 15-116)
MGKGRAPCCDKTKVGRGPWSHDEDLKLISFIHKNHENWRSPLKQAGLLRCGKSCRLRWI
NYLRPDVKRGNFSAEEEDTIIKLHQSFGNKWSKIASKLPGRTDNEIKNVWHTHLKKRLSS
ETNLNADEAGSKGSLNEEENSQESSPNASMSFAGSNISSKDDDAQISQMFHILTYSEFT
GMLQEVDPKPELLEMPFDLPDIWSFIDGSDSFQPPENRALQSESEDEVDKWFHLESELG
LEENDNQQQQHKQGTEDHSSSLESYELLIH*

>G1332 (1..606)
ATGGAATGCAAAAGAGAAGAAGGGAAGTCTTACGTGAAGAGAGGGTTGTGGAACAGAA
GAAGATATGATATTAAAAAGCTATGTTGAGACTCATGGTGAAGGAAACTGGGCAGACATT
TCTCGTAGATCCGGGTGAAGAGAGGAGGAAAAAGCTGTAGGCTGAGATGGAAGAAGTAT
CTAAGACCAATATCAAAAGAGGAAGCATGTCAACCAAGAACAAGACCTTATCATCCGC
ATGCATAAGCTTCTTGGAACAGATGGTTCGTTGATCGCTGGCTCCAGGTCGTACT
GACAATGAAGTGAAGAACTACTGGAATACTCATTTGAACAAGAAACCTAATCCCCGAAA
CAGAATGCACCTGAATCAATCGTCGGCGCCACTCCTTTACGGATAAGCCAGTTATGTCT
ACAGAACTGAGAAGAAGCCATGGAGAAGGAGGAGAAGAGGAGAGCAATACCTGGATGGAG
GAGACCAACCACTTTGGCTATGACGTCCACGTAGGATCTCCCTTGCCACTTATTTCCAC
TACCCAGACAACACTCTCGTGTGTTGACCCATGTTTTTCTTTACCGATTCTTTCTCTCTG
CTTTAG

>G1332 Amino Acid Sequence (conserved domain in AA coordinates:13-116)
MECKREEGKSYVKRGLWKPEEDMILKSYVETHGEGNWADISRRSGLKRGKSCRLRWKNY
LRPNIKRGSMSPPQEQDLIIRMHKLLGNRWSLIAGRLPGRTDNEVKNYWNTHLNKKPNRSK
QNAPEIVGATPFTDKPVMSTELRRSHGEGGEEESNTWMEETNHFYDVGVSPLPLISH
YPDNTLVFDPFCSFTDFPPL*

>G1334 (76..885)
ATAGCTCCCACTAATAGGAATCTCAAGCTTCTCACTCTCTCTGTTTTTCCATTGGACT
TTTGGAAACATAAGCTATGCAAACTGAGGAGCTTTTGTGCGCCACCACAGACTCCTTGGTGG
AATGCTTTTGGATCTCAGCCGTTGACTACAGAGAGCCTTTCCGGCGAAGCTTCTGATTCA
TTCACCGGAGTTAAGGCAGTTACTACGGAGGCAGAACAGGTGTGGTGGATAAACAACT
TCTACAACCTCTCTTCACTTTCTCACCTGGTGGTGAAGAAGAGTTCAAGAGATGTGCCAAG
CCTCATGTTGCTTTGCGGATGCAATCAGCTTGCTTCGAGTTTGGATTGTCTCAGCCAATG
ATGTACACAAAGCATCCTCATGTTGAACAATACTATGGAGTTGTTTCAGCATACGGATCT

CAGAGGTCTTCGGGCGGAGTAATGATTCCACTGAAGATGGAGACAGAAGAAGATGGTACC
ATCTATGTGAACCAAGCAGTACCATGGAATTATCAGGCGACGCCAGTCCCGAGCAAAG
GCTGAAAACTGAGTAGATGCCGTAAGCCATATATGCATCACTCAGGCCATCTCCATGCT
ATGCGCGCTCTAGAGGATCTGGCGGGCGTTTCTTGAACACCAAGACAGCTGATGCGGCT
AAGCAGTCTAAGCCGAGTAATTCTCAGAGTTCTGAAGTCTTTCATCCGAAAAATGAGACC
ATAAACTCATCGAGGGAAGCAAATGAGTCAAATCTCTCGGATTCTGCAGTTACAAGTATG
GATTACTTCTAAGTTCGTGGCTTATTCTCCTGGTGGCATGGTCATGCCTATCAAGTGG
AATGCAGCAGCAATGGATATTGGCTGCTGCAAACTTAATATATGATCAGCAGATAGGGGA
CAAGACATGATTGGTCACCAGTCCTTTGTCTTGTCCCTTATCTTTCAGCCAAACGGAAA
GAGAACTTGTCTTGGAAAAAAGACATTGAGTTTCCTTGGTTTATAAGATTGGTCCTTT
TACCATCCGTTTGGCTGTAAACAGGCAAATCATCTTGGCTCATGCTTCATCAAGTCTT
ATCTTCGTCTGTTTCTTCTACGCATCTTCATAAGATCTCTGAAGTAGTGAATAACATTT
CCTAGCATCATGTTTCAACTAGTGTGTGTGTGAAGAACTCTGCCTATTTCCAGATGAT
GTATTGTGTGTAACGTGTTTATGAAACAAACGTAAGACTTTCAAGTTAAAAA
AAAAAAAAAAAAA

>G1334 Amino Acid Sequence (domain in AA coordinates: 18-190)

MQTEBLLSPQTPWNNAFSGQPLTTESLSGEASDSFTGVKAVTTEAEQGVVDKQSTTLF
TFSPGGEKSSRDVVPKPHVAFAMQSACFEFGFAQPMYTKHPHVEQYGVVSAYGSQRSSG
RVMIPKMETEEDGTIYVNSKQYHGIIRRRQSRKAELSRCKPYMHHSRHLHAMRRPR
GSGGRFLNTKTADAQSKPSNSQSSEVFHPENETINSREANESNLSDSAVTSMDYFLS
SSAYSPGGMVMPKWNAAAMDIGCCKLNI*

>G1381 (32..802)

CAGCTTTAACTACTCTCTCTCTCTCAATGGGAAAAACAATCAACATAGAGAGTAG
TGCTACTCATCATCAAGACAATATTGTTTCCGTTATAACAGCCACGATATCCTCCTCCTC
CGTCGTAACGTTCTCGTCAGACTCTTGGTCTACCTCCAAAAGATCGTTAGTGCAAGACAA
TGACTCCGGAGGGAAACGGCGGAAGAGCAACGTTAGTGATGATAACAAGAAATCCGACGTC
GTATAGAGGAGTGAGGATGAGGAGTTGGGGAAAATGGGTGTCGGAGATTAGAGAGCCGAG
GAAGAAATCAAGAAATATGGCTTGGCACTTATCCAACGGCAGAGATGGCAGCTCGTGCTCA
TGATGTGGCGGCTTAGCTATTAAAGGCAACTCCGGTTTTCTTAATTTCCCTGAATTATC
CGGTTTGCTTCCCTCGTCCGTTAGCTGCTCTCCTAAGGATATAACAAGCTGCAGCTACCAA
AGCCGCGAAGCAACCAACGTCGACAAACCGGTTATCGATAAGAAATTAGCTGATGAGCT
AAGCCACTCTGAGTTGTTGTCTACCGCTCAGTCTTCGACTTCTAGTAGTTTCGTGTTTTCT
TTCGGACACGTCGGAGACTTCTAGTACGGACAAGGAAAGCAACGAAGAGACGGTGTGTTGA
TTTGCCGGACCTTTTACGGACGGGCTTATGAACCCAAACGATGCGTTTTGTTTATGCAA
CGGCACCTTTACGTGGCAGCTTTACGGAGAGGAGGATGTAGGGTTCAGGTTTGAAGAGCC
GTTTAATTGGCAAAATGACTAAACCGCCCTCCACTTGCTTACTGTAATTACTAACATATA
ATTTTCTTGATAAAGAACATATATTTCCATTACGGTATTAACTAATCTTTTCTATCCTTT
TCTCTTTCTTGTTTCTACATCTGAGTATATTGTCACTATGTGAAAAAATTGATCTCGTT
TTGAATATTCTTTTCAAAATTGAAGTAACGCAAGTGATTGATAAAAAAAAAAAAAA

>G1381 Amino Acid Sequence (domain in AA coordinates: TBD)

MGKQINIESSATHHQDNIVSVITATISSSSVVTSSSDSWSTSKRSLVQDNDSSGKRRKSN
VSDDNKNPTSIRGVMRMSWGKWWSEIREPRKKSRIWLGYPTAEMAARAHDAALAIKGN
SGFLNFPPELGLLRPVSCSPKDIQAAATKAAEATTWHKPVIDKKLADELHSELLSTAQ
SSTSSSFVFSSTSETSDTKESNEETVFDLPDLFTDGLMNPNDADFCLCNGTFTWQLYGE
EDVGFRFEPEPFNWQND*

>G1382 (90..1763)

CTCTCATTTCGCCATAGCTGAGAGCTTCTTCTACTTTCCCTTAGCTTCTTTTTTCCTTCA
TTTTTGTCTACCCFTGCGAATCTCTGAAATGAACCTCAAGCTAATGACCGGAAGGAGT
TTCAGGGAGATTGTTTCGGCGACGGGAGATCTCACGGCAAAGCACGATTTCAGCTGGAGGAA
ACGGAGGTGGAGGTGCTAGGTATAAGCTGATGTACCCGGCCAAGCTTCCGATCTCGAGGT
CGACTGATATCAGATTCTCTGGGTGAGTCCGACTTCGTTTTTGGAAATCTCCTGTTT
TCATCTCCAACATCAAGCCAGAACCCTCCCTACTACTGGTCTTTGTTCAAGCCTCGAC
CAGTGCAATTTCTGTAGCTCAAGTTCTTATACAGGCAGGGGGTTCCATCAGAACACCT
TTACTGAGCAGAAGTCCAGTGAATTTGAGTTCAGACCTCCTGCATCAAATATGGTATATG
CAGAGCTTGGCAAGATTAGAAGTGAGCCACCAGTACATTTTCAAGGCCAGGGCCATGGAT
CCTCACACTACCTTCTTCGATCAGTGATGCTGCAGGTTCTCAAGTGAGCTAAGCCGGC
CAACTCCTCCTTGTCAGATGACACCAACGAGCTCAGATATTCCGGCTGGATCTGATCAAG

AGGAATCAATCCAGACTTCCCAAAATGACTCCAGAGGAAGCACTCCATCCATCTTGGCTG
ATGATGGTTATAACTGGAGAAAATATGGTCAAAAGCATGTCAAAGGGAGTGAATTTCCCC
GGAGCTATTATAAATGTACACATCCTAATTGTGAAGTGAAAAAGTTATTTGAAAGATCTC
ATGATGGGCAGATCACCAGATATTATATACAAGGGTACACATGACCATCCTAAACCTCAAC
CTGGTCGCCGAACTCTGGTGGTATGGCTGCACAAGAAGAAAGGCTAGACAAGTATCCTT
CTTCAACTGGCCGAGATGAGAAGGGATCTGGCGTCTACAACCTGTCTAACCCCAATGAAC
AACTGGTAACCCTGAAGTACCTCCTATCTCAGCATCTGACGATGGTGGAGAAGCGGCAG
CGTCAAATAGGAATAAAGATGAGCCGGACGATGATGATCCATTCTCAAAACGGAGGAGGA
TGGAGGGTGCGATGGAATAAATCCACTAGTGAAACCCATCCGGGAGCCTCGGGTTGTTG
TTCAAACCTCTGAGTGAAGTTGACATCTCTGGATGATGGTTATAGATGGCGCAAAATATGGGC
AGAAAGTCGTAAGGGGGAACCCAAATCCCAGGAGCTACTACAAATGCACAGCTCATGGAT
GCCAGTGAGAAAACACGTGGAGAGAGCATCACATGATCCAAAGCTGTAATAACAACAT
ACGAAGGCAAAACAGATCATGATGTTCCCACTTCAAAGTCTAGCAGCAATCACGAAATCC
AGCCTCGGTTCCAGACAGATGAAACAGACACCATCAGCCTCAATCTTGGTGTGGAATCT
CATCTGATGGACCTAACCACGCTTCCAACGAACATCAGCACCAGAATCAACAACCTGTCA
ACCAAACCTCACCCAAATGGAGTCAATTTTCAAGTTTGTTCATGCTAGTCCCATGTCATCCT
ACTATGCTAGCTTAAATAGCGGTATGAATCAGTACGGCCAGAGAGAAACAAAGAACGAGA
CTCAAAATGGTGACATCTCGTCTTGAACAATTCATCTTACCCATATCCGCCCAACATGG
GGAGAGTACAATCGGGTCCGTAAAAACAAAAGTAAGCAACATTATGTACGGGATCTTCTT
AGGTTAGGAATGGGACGAGGCTTGTCTATATAATTCTTATTTCTTACAGAGAGCTGA
TCTTGATTCAAATATCTCCACCATATATATTGTTTGTGTCACCTGTATTGAGTTCCAA
AAATGTTATGTAAAAATACACAACAAGATGTTAATGCTTTTATTTAAACAAGAAACAGCA
ATATTACTACAAAAAATAAAAAAAAAA
>G1382 Amino Acid Sequence (domain in AA coordinates: 210-266, 385-437)
MNPQANDRKEFQGDSCATGDLTAKHDSAGNGGGGGARYKLMSPAKLPISRSTDITIPPL
SPTSFLESPVFIISNIKPEPSPTTGLFKPRPVHISASSSSYTGRGFHQNTFTEQKSSEFE
FRPPASNMVYAEELGKIRSEPPVHFQGGHSSHPSSISDAAGSSSELSPRTPPCQMTPT
SSDIPAGSDQEEISIQTSONDSRGSTPSILADDGYNWRKYQGKHVKGSEFPRSYYKCTHPN
CEVKKLFERSHDGQITDIIYKGTGHDHPKQPGRNRNSGMAAQEERLDKYPSSSTRDEKGS
GVYNLSNPNEQTNPEVPPISASDDGGEAAASNRNKDEPDDDDPFSKRRRMEGAMEITPL
VKPIREPRVVVQTLSEVDILDDGYRWRKYQGKVVGRGNPNRSYYKCTAHGCPVRKHVERA
SHDPKAVITTYEGKHDHVDPTSKSSSNHEIQPRFRPDETDITISLNLGVGISSDGNPHASN
EHQHQNQLVNTQTHPNGVNFRFVHASPMSSYYASLNSGMNQYQRETNETQNGDISSLN
NSSYPYPPNMGRVQSGP*
>G1435 (8..904)
GTGAAACATGGGGAAGGAAGTTATGGTGAGCGATTACGGTGACGACGAGAGAAGACGC
CGGCGGCGCGATGAATATAGGATTCCGGAATGGGAAATTGGTTTACCCAACGGAGATGA
TTTGACTCCGTTATCTCAATATCTAGTCCCGTCGATTCTCGCGTTAGCTTTCAGCATGAT
CCGAGAACGAAGCCGTACAATTCACGACGTCAATCGCGCGTCGCAATCACGCTCTCTTC
GTTGAGAAGCAGTACCAATGCTTCGTCTGTGATGGAGGAGGTCGTGGATCGAGTTGAATC
GAGTGTTCAGGATCAGATCCGAAGAAACAGAAGAAATCGGATGGTGGTGAAGCAGCGGC
GGTGGAGGATTCCACGGCGGAGGAAGGAGACTCCGGGCCTGAAGACGCGTCTGGGAAGAC
ATCGAAACGACCGGTTTAGTGTGGACACCGCAGCTACACAAGAGATTGTGGACGTTGT
GGCTCATCTAGGGATTAAAAACGCAAGTCCGAAGACGATTATGCAGCTGATGAACGTGGA
AGGACTTACTCGTGAGAAGCTTGCCTCTCATTTGCAGAAATATAGGCTTTACCTTAAACG
GATTCAAGGATTGACGACGGAAGAAGATCCTTATTCTGTCGCGATCAGCTCTTCTCTTC
AACGCCGGTTCTCCACAGAGCTTTCAAGACGGCGGAGGAAGTAACGGAAGTTGGGGGT
TCCGGTTCCGGTTCEGTGATGGTGCCTATTCCAGGCTATGGGAATCAAATGGGTATGCA
AGGATATTATCAACAGTATAGTAACCATGGCAATGAATCAAACCAATATATGATGCAGCA
GAATAAGTTTGAACAATGGTGACATATCCTTCTGTTGGTGGTGGTGACGTGAATGACAA
GTAAATGGATCTTAAAGGTCTATAATTTGCTCTACAGAGAGATACTGGTTCTTGGCTTAT
GGTTTATTTTCCCACTTCATGAGGTTGTTGTGACTTTTAATTCTCCATGTTTTCCACACA
AGTCTTTATTGCTTTTGTATAGAAAATGATTTGAGAAAATCACTGGGAAGCTTGGTATT
GTTGGAGGATGAAGCCTTCTATGAATGATTTAGTTTCTACTGTCTCCATTCTTTATGAG
GTAATAAAGCCTTCTTTGCTCATCGCTTGTAGTCTTCTTAAATTCAAGACAGCGTCACA
TGTTTGTTCGGTTATGTTAATTGTTTCTTTCTTTGGATAATGAAGATAGCATCAGGTCTC
ATGTCTCCTCACTTTGATAAA

>G1435 Amino Acid Sequence (domain in AA coordinates: 146-194)
MGKEVMVSDYGDGDDGDEYRIPEWEIGLPNGDDLTPLSQYLVP SILALAFSMIPE
RSRTIHDVNRASQITLSSLRSSTNASSVMEEVVDREVSVPGSDPKKQKSDGGEAAAVE
DSTAEEGDSGPEDASGKTSKRPLRVWTPQLHKRFVDVVAHLGIKNAVPKTIMQLMNVEGL
TRENVAHLQKYRLYLKRIQGLTTEEDPYSSSDQLFSSTPVPPQSFQDGGGSGNKLGVVPV
PVPSMVPPIPGYGNQMGMQGYQQYQYNSHGNESNQYMMQONKFGTMVTPSVVGGGDVNDK*

>G1537 (1..783)

ATGGAACCGAAGTAAACCGCAGGAACAGCAAGCAGTTCAAGATGGAACCCAACGAAAGAT
CAGATCACGCTACTGGAAAATCTTTACAAGGAAGGAATACGAACCTCCGAGCGCCGATCAG
ATTACAGCAGATCACCGGTAGGCTTCGTGCGTACGGCCATATCGAAGGTAAAAACGCTCTT
TACTGGTTCAGAACCATTAAGGCTAGGCAACGCCAAAAGCAGAAACAGGAGCGCATGGCT
TACTTCAATCGCCTCCTCCACAAAACCTCCCGTTTCTTCTACCCCCCTCCTTGCTCAAAC
GTGGGTTGTGTGCTAGTCCGTACTATTTACAGCAAGCAAGTGATCATATGAATCAACAT
GGAAGTGTATACACAAACGATCTTCTTACAGAAACAATGTGATGATTCCAAGTGGTGGC
TACGAGAAACGGACAGTCAACAACATCAGAAACAACCTTTCAGACATAAGAACAACAGCA
GCCACAAGAAATGCCAATTCTCCGAGTTCAGTCAAGTTTGACAGATTGCCCCCTCCGTGAT
AACTGTTATGCCGCTAGGACATTAACGCTCAATTCCAGTGGACGGAACAACTCCCTCTT
TTTCTCTTTCAGCCTTTTGAATGCAAGTAATGCTGATGGTATGGGAAGTTCCAGTTTGGC
CTTGGTAGTGATTCTCCGGTGGATTGTTCTAGCGATGGAGCCGGCCGAGAGCAGCCGTTT
ATTGATTTCTTTTCTGGTGGTCTACTTCTACTCGTTTCGATAGTAATGGTAATGGGTG
TAA

>G1537 Amino Acid Sequence (domain in AA coordinates: 14-74)
MENEVNAGTASSSRWNPTKDQITLLENLYKEGIRTPSADQIQITGRLRAYGHIEGKNVF
YWFQNHKARQRQKQKQERMAYFNRLHLKTSRFFYPPPCSNVGCVSPYYLQOASDHMMNQH
GSVYTNDDLHNRNMIPSGGYEKRTVTQHQQLSDIRTTAATRMPISPSLRFDRFALRD
NCYAGEDINVNSSGRKTLPLFPLQPLNASNADGMGSSSFALGSDSPVDCSSDGAGREQPF
IDFFSGGSTSTRFDSNNGNL*

>G1545 (67..729)

CATCACCAATCTTTTGAATCTAAGAGAGAGAAGAAGAAGGTCTAGAGAACGAAAAGA
AGAAACATGAATAACCGAATGTAGATGATCATAATCTTCTACTCATTTCTCAATTGTAC
CCTAATGTCTATACTCCATTAGTACCACAACAAGGAGGAGAAGCAAAACCAACACGGCGG
AGGAAAAGGAAGAGCAAGAGTGTGTGGTGGCAGAGGAGGGTGAAAACGAAGGCAATGGG
TGGTTTAGAAAGAGAAAATTGAGTGATGAGCAAGTAAGAATGTTGGAGATTAGCTTTGAA
GACGATCATAAGCTTGAATCCGAGAGGAAAGATCGGCTTGCTTCTGAGTTAGGGCTTGAT
CCTCGTCAAGTCGCCGTCTGGTTCCAAAACCGCGTGACGGTGAAGAACAACAGAGTC
GAGGATGAATACACTAAACTCAAGATGCATACGAAACCACCGTCGTTGAGAAATGTCTGT
CTTGATTCTGAGGTTATTACCTAAAGGAACAACCTTTACGAGGCTGAAAGAGAGATCCAA
CGGCTTGCAAAAAGAGTTGAAGGAACCTTTAAGTAACAGTCCTATCTCATCTCTGTGACC
ATTGAAGCCAATCATACGACACCGTTTTTTTGGAGATTACGACATCGGATTTGACGGTGAG
GCTGACGAGAAGTGTCTACTCGCCAGATTACATTGATGGATTAGACTGGATGAGCCAA
TTTATGTAAAAAACTATAAGCTAATCTATTTTTCAGTCGTAGTATAG

>G1545 Amino Acid Sequence (domain in AA coordinates: 54-117)
MNNQNVDDHNNLLISQLYPNVYTPLVPQQGGEAKPTRRRKRKSKSVVVAEEGENEGNGWF
RKRKLSDEQVRMLEISFEDDHKLESEKDRLLASELGLDPRQVAVWFQNRARWKNKRVED
EYTKLKNAYETTVEKCRDLSEVIHLKEQLYEAREIQRLAKRVEGTLSNSPISSSVTIE
ANHPTTFPGDYDIGFDGEADENLLYSPDYIDGLDWMSQFM*

>G1641 (1..867)

ATGGAGGTTATGAGACCGTCGACGTACACGTGTCAGGTGGGAAGTGGCTCATGGAGGAA
ACTAAGAGCGGCGTCGAGCTTCTGGTGAAGGTGCCACGTGGACGGCGGCAGAGAACAAG
GCATTCGAGAATGCTTTGGCGGTTTACGACGACAACACTCCTGATCGGTGGCAGAAGGTG
GCTGCGGTGATTCCGGGGAAGACAGTGAGTGACGTAATTAGACAGTATAACGATTGGAA
GCTGATGTCAGCAGCATCGAGGCCGTTTAAATCCCGGTCGCCGTTACATCACCTCGCCG
CCTTTCACTCTAGATTGGCCCGGCGCGGTGGCGGATGTAACGGGTTTAAACCGGTCAT
CAGGTTTGTAAATAAACGGTCGAGGCCGGTAGATCGCCGGAGCTGGAGCGGAAGAAAGGC
GTTCTTGGACGGAGGAAGAACAAGCTATTTCTAATGGGTTTGAAGAAATATGGGAAA
GGAGATTGGAGAAACATATCTCGGAACCTTGTGATAACGCGAACGCAACACAAGTAGCT
AGCCACGCCCAAAAGTACTTCATCCGGCAACTTCCGGCGGCAAGGACAAGAGACGAGCA

AGCATTACGACATAACCACCGTAAATCTCGAAGAGGAGGCTTCTTTGGAGACCAATAAG
AGCTCCATTGTTGTTGGAGATCAGCGTTCAAGGCTAACCGCGTTTCCTTGGAAACCAAACG
GACAACAATGGAACACAGGCAGACGCTTCAATATAACGATTGGAAACGCTATTAGTGGC
GTTTCATTACATACGGCCAGGTTATGATTGGAGGGTATAACAATGCAGATTCTTGCTATGAC
GCCCCAAACACAATGTTTCAACTATAG

>G1641 Amino Acid Sequence (domain in AA coordinates: 139-200)
MEVMRPSTSHVSGGNWLMEEETKSGVAASGEGATWTAENKAFENALAVYDDNTPDRWQKV
AAVIPGKTVSDVIRQYNDLEADVSSIEAGLIPVPGYITSPFFTLWDAGGGGGCNGFKPGH
QVCNKRSQLAGRSPELERKKGVPTWEEHKLFLMGLKKYKGDWRNISRNFVITRPTQVA
SHAQKYFIRQLSGGKDKRRASIHDIITVNLEEEASLETNKSSIVVDQRSRLTAFPNWQT
DNNGTQADAFNITIGNAISGVHSYGQVMIGGYNNADSCYDAQNTMFQL*

>G165 (19..699)
CTTCAAAACATCTAAAAAATGGTGAAAAAACTCTTGGTCGTAGAAAGGTAGAGATAGTG
AAAATGACTAAGGAATCAAAACCTTCAAGTCACATTTTCCAAGAGAAAAGCTGGTCTTTTT
AAGAAGGCTAGTGAATTTTGACATTTATGTGATGCAAAAATTCGATGATCGTGTTTTCA
CCAGCTGGAAGTATTTTCTTTTGGTCATCCAAATGTTGATGTTCTGCTTGACCACTTT
CGAGGGTGTGTTGTAGGACACAACAACACAACCTTGATGAAAGCTACACAAAGCTTCAT
GTTCAAAATGCTCAACAAATCCTACACTGAGGTGAAGGCGGAAGTAGAAAAAGAACAAAAG
AATAAGCAGTCGCGGGCTCAAAATGAAAGAGAAAACGAAAACGCTGAGGAGTGGTGGAGT
AAGTCTCCATTAGAACTCAACTTAAGTCAATCAACCTGTATGATACGTGTTCTTAAAGAT
TTGAAGAAGATAGTTGATGAAAAAGCAATTCAATTAATCCATCAAACAAACCCAACTTC
TATGTTGGAAGTTCTAGCAATGCTGCTGCTCCAGCAACTGTTAGTGGTGGTAATATCTCC
ACAAACCAGGGGTTCTTTGATCAAAACGGAATGACGACTAATCCTACTCAAACACTTCTG
TTTGGATTGATATTATGAATCGCACACCAGGAGTTTAAATAAGTCTATCCTCATTATGG
GTCTTGGTACTATAAGTTCATCTCTCTCGTTGTTGACTTTTAAAGTCTCCAATAGTTTGT
TGTTG

>G165 Amino Acid Sequence (conserved domain in AA coordinates: 7-62)
MVKKTLGRRKVEIVKMTKESNLQVTFSKRKAGLFKKASEFCTLCDAKIAMIVFSPAGKVF
SFGHPNVDVLLDHFRCVVGHNNTNLDES YTKLHVQMLNKS YTEVKA EVEKEQKNKQSRA
QNERENENAE EWWSKSPLELNLQSQSTCMIRVLKDLKKIVDEKAIQLIHQTNP NFYVGS
SS NAAAPATVSGGNISTNQGF FDQNGMTTNPQTLLFGFDIMNRTPGV*

>G1652 (77..1078)
AGCAAGTCCAAATCTCCCTCTCTCTCTCTATCTCTCTATAGAAGATTTTTTAAAC
TAAGAAGCTAGCGATAGGCCACAGCGATGAACGTTTTCTCTACCAAATGGTCTCCGA
ATTGGATATAGAAGAAATATAGTATCATCCCAATTCACATGAACCTACTCGTCGGAGA
TGTTCCACAGTCTCTCTCTCTCTGATGATACCACCACTTGTTATAACCTTGATGCTTC
TTGTAATAAAAGTTTGGTAGAAGAAAGACCTTCAAAGATCCTCAAGACCACTCACATATC
ACCAAACCTTACATCCTTTTTCTTCTTCTAATCCTCCTCCTCAAAGCACCAGCCCTCTTC
TAGGATTTCTTTCTTTTGAAGACAGGTTTACATGTTATGAATCACAACCTCTCCAAACTT
AATATTTAGCCCCAAGGACGAAGAAATTGGATTACCAGAGCATAAGAAAGCCGAGCTGAT
AATAAGAGGGACAAAGAGAGCTCAATCCTTGACTCGAAGCCAATCAAATGCTCAAGATCA
CATACTGGCAGAGAGAAAACGGAGAGAGAGAAGCTTACTCAAAGATTGTAGCTCTTCCGC
GCTAATTCCTGGCCTAAAGAAGATGGACAAGGCTTCTGTGTTGGGAGATGCAATAAAGCA
TATAAAGTACCTCCAAGAGAGTGTGAAAGAGTATGAGGAACAAAAGAAGGAAAAGACAAT
GGAATCAGTGGTTCTTGTAAAGAAGTCTAGTCTGGTTTTAGATGAAAATCATCAACCATC
ATCATCATCTTCCTCAGATGGAAATCGCAATAGCTCGAGCTCAAATCTTCAGAAATAGA
AGTTAGGGTTTCAGGAAAAGATGTTCTTATTAAGATCCTATGCGAGAAGCAAAAGGGTAA
TGTGATCAAGATTATGGGGGAGATTGAAAAGCTTGGTTTGTCTATCAACACAGCAATGT
CTTGCCCTTTTGACCCACTTTTGACATCTCTATTATCGCTCAGAAGAATAACAATTTGA
TATGAAAATCGAGGATGTTGTGAAGAACTTGAGTTTGGCTTATCAAAGCTCACTTAATT
GGTTTCACGTACATACATATACACATTCATCATCGATTTCTCCGATCGAAGAATCCAAA
ATCAGTTTTTCCATGAAAGTGGTTTTTTAGTTGTTAAGTTTGTGTATGGAGATCTTAA
GTCATTTAAAGATCCTTGTTCTTGTTGTTGTTAAGTGTGCTTTAAGATGCATATCATCAA
TGTTTAGTAATTTATTTCTCCAGTTTCATTTGGGACGGAATTTTTTTCGAGTTGTTGG
ATATATATTTCTCGATGTAAAGCATTTCGTTAGTTTAAATAAACGTCCGATATGTTTCT
TTGAAAA

>G1652 Amino Acid Sequence (domain in AA coordinates: 143-215)

MATAMNVFSTKWSSELDIEEYSIIHQFHMNSLVGDVPQSLSSLDTTTTCYNLDA SCNKS L
VEERPSKILKTTTHISPNLHPFSSSNPPPKHQPSRILSFEXTGLHVMNHNSPNLI FSPK
DEEIGLPEHKKAEIIRGTRKRAQSLTRSQSNQDHLAERKRREKLTQRFVALSALI PGL
KKMDKASVLGDAIKHIKYLQESVKEYEEQKKEKTMESVVLVKSSSLVDENHQPS SSSSS
DGNRNSSSSNLPEIEVRVSGKDVLIKILCEKQKGNVIKIMGEIEKLGLSITNSNVL PFGP
TFDISIIAQKNNNPFMDKIEDVVKNLSFGLSKLT*

>G1655 (132..755)

TTTCTAACTAGTCACATTGAGAGAGAGAGAGAGAGAGAGAGAAAGAGAGACTCTCAGAATCTGAAG
AAGAAGAAGAGATTGTTGTTTTTGCCTTTTATCATCGGTTTCTTTGAATCTCTGGTTTTA
AATCGGATTTAATGGTGGAGTCTCTGTTCCCGAGCATCGAAAACACAGGTGAATCGTCTC
GAAGAAAGAAGCCGAGGATATCAGAGACGGCGGAGCGGAGATAGAGGCACGACGTGTCA
ACGAAGAAAGCTTGAAGAGATGGAACGAATCGTGTGCAACAGATCTACGCTTGTAAAGC
TCGTGCAAGCTTTACGCCGAGTTTCGTGAGAGATCTCCACCACCAGCAACAACGAGACCG
ATAAATCGTCTCCGGCGCGGCGAGGGAGATACGTGATACGGCGGATCGAGTTCTAGCTG
CGTCCGCTCGTGGTACGACTCGGTGGAGCAGAGCGATTTTAGCGAGTCGCGTCCGAGCGA
AGCTGAAGAAACATAGAAAGGCGAAAAAGTCAACGGGAAATTGTAAATCGAGAAAAGGTC
TCACGGAGACGAATCGGATTAAGTTACCGGCGGTTGAGAGAAACTGAAGATTCTTGCC
GTTTGGTTCTGGTTGCCGGAAGTCTCTGTACCGAATCTTTTAGATGAAGCGACCGATT
ACATCGCAGCGTTAGAGATGCAGGTTTCGAGCCATGGAGGCTCTCGCCGAACTTTTAACCG
CAGCCGCACCACGGACGACGTTGACCGGAACCTTAACGGCGGCAGTTAGTTTGTCAAGTGT
TAATTAGCTTTTCTTTTACCTTTTTTACCCTTTTATTTTGGCTTCAAGTGTTTTTTTTTTC
TCGTGACGCGATTTTAATTTATTAATTTCA

>G1655 Amino Acid Sequence (domain in AA coordinates: 134-192)

MVESLFPSIENTGESSRRKKPRISETAEAEIEARRVNEESLKRWKTRVQQIYACKLVEA
LRRVRQRSSTTSNNETDKLVSGAAREIRDADRVLAAARGTTRWSRAILASRVRAKLKK
HRKAKKSTGNCKSRKGLTETNRIKLPAVERKLKILGRLVPGCRKVSVPNLLDEATDYIAA
LEMQVRAMEALAEELLTAAAPRTTLTGT*

>G1671 (188..751)

TCCCACTATCCTTCGCAAGACCCTTCTCTATATAAGGAAGTTCATTTCAATTTGGAGAGG
ACACGCTGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACCCTCTCTATATAATCTTC
TTCTACACACACACACACACGCAACCATATACGTACATGTGAAGTAGTGAGATCAATATC
GTTAGCAATGAATCTACCACCGGGATTAGGTTTTTTCCGACCGATGAAGAGCTCGTCGT
TCATTTCTCTCCACCGGAAAGCTTCCCTCTTGCTTGTACCCCTGATGTATCCCCGACCT
TGATCTTTACCATTACGATCCTTGGGACCTTCCCGGAAAGCTTTGGGAGAAGGGAGGCA
ATGGTACTTCTATAGTAGAAAGACACAAGAGAGAGTGACAAGCAATGGGTATTTGGGGATC
AATGGGAATGGACGAGCCAATCTACACAAGCTCCACACACAAGAAAGTGGGAATCAAAAA
GTATCTAACTTTCTATCTCGGAGATTCTCAGACTAATTGGATCATGCAAGAAATATCCCT
CCCGGATTCTCTTCTCATCTAGTCGATCTTCTAAGAGATCAAGCCGTGCTTCTAGTTC
TAGTCACAAACCGATTATAGCAAGTGGGTGATATGCAGAGTGTATGAGCAAAATTGCAG
TGAGGAGGAAGACGATGATGGGACAGAACTCTCATGTTTGGATGAAGTGTTTTTGTCTTT
AGATGATCTTGACCAAGTAAGCTTACCGTAATAAAGACAGAAGCACCCAAGAAGAGAAAA
AAAAAAAAGGGTTTAGTGGGCAATTATTTCTAAGCGACCGCTCTAGACAGGCCTAGTAC
CGGATCCTCTAGCTAGAGCTTTCGTTTCGTATCATCGGTTTCGACAACGTTTCGTCAAGT

>G1671 Amino Acid Sequence (domain in AA coordinates: TBD)

MNLPPGFRFFPTDEELVVHFLHRKASLLPCHPDVIPDLDLHYHDPWDLPGKALGEGRQWY
FYSRKTQERVTSNGYWGSMGMDPIYTSSTHKKVGIKKYLTFLGDSQTNWIMQEYSLPD
SSSSSRSSSRSSSRSSSKPDYSKWVICRVYEQNCSEEDDDGTELSCLDEVFLSLDD
LDEVSLP*

>G1756 (71..1003)

ATATGTACTTGTACACCAACCCACCAAAAGAGATAAAAGAGGAAACAAAACCTCGAAAAG
AGAGAGATATATGGGTGAGGTGGCTTATATGGACGAAGGAGACCTAGAAGCAATAGTCAG
AGGCTACTCCGGCTCCGGAGACGCGTTTTCCGGCGAAAGTTCCGGTACGTTTTACCTTC
GTTTTGCCTACCGATGGAGACGTCTAGTTTCTACGAACCGGAGATGGAGACAAGTGGCTT
AGATGAGCTCGGTGAACCTTTACAAACCTTTTACCCTTTCTCCACACAAACGATCCTCAC
AAGCTCGGTCTCTCTCCCTGAAGATTCAAAACCTTTCCGAGATGACAAGAAACAACGATC
ACATGGTTGTCTTTTATCCAACGGATCAAGAGCTGATCATATCCGAATTTCAGAATCCAA
ATCAAAGAAAAGCAAGAAGAATCAACAGAAGAGAGTTGTTGAGCAAGTGAAAGAAGAGAA

TCTGTTGTCGGACGCATGGGCGTGGCGTAAATACGGGCAGAAACCCATCAAAGGATCTCC
ATACCCAAGGAGTTATTACAGATGCAGTAGCTCAAAGGGTGTGGCAAGAAAACAAGT
CGAAAGAAATCCTCAAACCCGGAGAAATTCACCATAACATACTAATGAGCACAAATCA
TGAAC TACCAACCCGGAGAACTCATTAGCCGGTTCGACTCGAGCAAAAACCTCCCAACC
CAAACCAACCTTAACCAAAAAATCCGAAAAAGAGTTGTTTCTTCCCCTACAAGTAATCC
TATGATCCCATCCGCTGATGAATCTTCTGTTGCGGTTCAAGAAATGAGCGTTGCGGAAAC
GAGTACGCACCAAGCGGCTGGAGCAATCGAGGGCCGCGCTTGAGTAACGGTTTACCATC
GGATTTGATGTCCGGGAGCGGAAC TTTCCAAGTTTACC GG TGACTTCGATGAAC TATT
GAATAGCCAAGAGTTCTTCAGTGGGTATTTATGGAATTACTAGAGAGCATTAGGTGTATG
TATATATATAT

>G1756 Amino Acid Sequence (domain in AA coordinates: TBD)
MGEVAYMDEGDLEAIVRGYS GSGDAFSGESSGTFSPSFCLPMETSSFYEP EMETSGLDL
GELYKPFYFSTQTILTSSVSLPEDSKPFRDDKKQRSHGCLLSNGSRADHIRISESKSKK
SKKNQQRVVEQVKEENLLSDAWARKYQKPKIKGSPYPRSYRCSKGLARKQVERN
PONPEKFTITYTNEHNHELPTRRNSLAGSTRAKTSQPKPTLTKKSEKEVVSSPTSNPMIP
SADESSVAVQEMSVAETSTHQAAGAI EGRRLSNGLPSDLSGSGTFPSFTGDFDELLNSQ
EFFSGYLWNY*

>G1757 (250..1224)

ATCACCAATCCTATAACACTCTCATTCTCATCATATCATTCTTCAATCTATATAACCCAT
TCTTAATTATACTCAACACACATTATATTTTCTGATCATATCATTCTTTCAGTCCATCT
ATATAACCAATTCTTGATTTTACTTAAACACACATTATACATCTTCTCATCATAGTT
TGTATCAATTTCTTAGAGTAACTACCTAAAGGAAAAAAATCTATTTTGGGAATCAT
ATACTAAAAATGGAAGGAAGAGATATGTTAAGTTGGGAGCAAAAGACATTGCTAAGCGAG
CTTATCAATGGATTGATGCGGCCAAAAAGCTTCAGGCACGACTTAGAGAAGCTCCGTCG
CCGTCGTCATCATTTTTCATCACC GCGACGCTGTTGCTGAGACTAACGAGATTCTGGTG
AAGCAGATAGTTTCTTCTACGAGAGATCTCTTCTTCTGCTAAACTGGTCATCTCACC
AGCGTACAACCTTATCCGACGCCGTTACTGTAGTCCCGGTGGCAAATCCCGGCAGTGTT
CCAGAATCTCCGGCATCGATAAACGGAAGTCCGAGAAGTGAAGAGTTTGCCGATGGAGGA
GGTTCTAGCGAGAGTCATCATCGCCAAGATTACATTTTCAATTCAAAGAAAAGAAAGATG
TTACCAAAGTGGTCAGAAAAAGTGAGAATAAGCCCAGAGAGAGGCTTAGAAGGACCTCAA
GATGATGTCTTTAGCTGGAGAAAATATGGTCAAAAAGACATTTTAGGCGCAAATTTCCCA
AGGAGTTATTACAGATGCACACATCGTAGCACACAAAACCTGTTGGGCAACGAAACAAGTC
CAGAGATCAGACGGGGATGCTACGGTTTTCGAAGTGACGTACAGAGGAACACACACTTGT
TCGCAGGCGATCACAAGAACACCACCATTAGCCTCGCCGGAGAGCGACAAGACACCAGA
GTCAAACCAGCCATTACCCAAAAGCCAAAGGATATTCTCGAGAGTCTTAAATCCAACCTTA
ACCGTTTGAACCGATGGGCTTGATGATGGTAAAGACGTTTTCTCGTTCCCTGATACGCCG
CCGTTTTACAAATTACGGAATATCAACGGCGAGTTCCGCCACGTGGAGAGTTCTCCGATC
TTCGACGTTGTTGACTGGTTCAATCCAACGGTCGAGATTGACACAACCTTCCCCCGGTTT
TTACACGAGTCGATTTATTATTAATTAATAATTTGTAACAGAGAAATAGATAGTAAGTAGT
AAGTAATGATCAGCGAGAGTTAAACATAAAAGTACTTAGAGTAATCTAACGATGCATAA
TAAGGAATGTTCAACAGGACTTGAACATGATTTCAATACTAAGAGAGATTTATCTAGCTA
CTGGTAGTAGCCGAGACTTCTTGTGTAGCTTCACTTNCCTTTTGTGCTT

>G1757 Amino Acid Sequence (domain in AA coordinates: 158-218)
MEGRDMLSWBQKTL LSELINGFDAAKLQARLREAPSPSSSFSPATAVAETNEILVKQI
VSSYERSLLLLNWSSSPVQLIPTVTVVPVANPGSVPE SPASINGS PRSEEFADGGGSS
ESHHRQDIYFNSKKRMLPKWSEKVRISPERGLEGPQDDVFSWRKYQKDILGAKFPRSY
YRCTHRSTQNCWATKQVQRSDG DATVFEV TYRGTH TCSQA ITRTPPLASPEKRQDTRVKP
AITQPKPDILES LKSNLTVR TDGLDDGKDVFSFPDTPPFYNYGTINGEFGHVESSPIFDV
VDWFNPTVEIDTTFPAFLHESIYY*

>G1782 (1..927)

ATGCAAGTGTTC AAAGGAAAGAAGATTCATCTGGGGAAACTCAATGCCTACAACAAAT
TCAAATATTCAAGGATCTGAATCTTTCAGCTTGACTAAGGATATGATAATGTCTACAACA
CAATTACCCGCGATGAAACATTCGGGTTTGAGCTGCAAAATCAAGATTCAACCTCATCA
CAATCTACTGAAGAAGAATCAGGCGCGGTGAAGTTGCAAGCTTTGGAGAATATAAGCGT
TATGGATGCAGCATTTGTTAATAACAATCTCTCAGGTTACATCGAAAACCTGGGAAAGCCT
ATTGAAAATTATACTAAGTCAATTACTACCTCGTCGATGGTGCTCTCAAGACTCTGTGTTT
CCTGCTCCTACTTCTGGTCAAATATCTTGGTCTCTTCAATGTGCTGAAACGTCACATTTT

AATGGTTTCTTGGCTCCTGAATATGCATCAACACCAACGGCGCTGCCACATTTAGAGATG
ATGGGTTTGGTTTCTTCAAGAGTGCCATTGCCTCATCACATTCAAGAGAATGAACCAATA
TTTGTCAATGCGAAACAGTATCATGCGATTCTCCGTCGCAGGAAGCACCGTGCTAAACTC
GAAGCTCAGAACAACTCATCAAATGCCGTAAACCGTACCTTCATGAGTCTCGCCATCTT
CATGCTTTAAAGAGAGCTAGAGGCTCCGGTGGACGTTTCTCAATACAAAGAAGCTTCAA
GAATCATCAAACCTACTGTGTTCTTCTCAAATGGCAAATGGACAAAATTTCTCTATGAGC
CCTCACGGTGGTGGGAAGCGGAATCGGGTCTAGTTCGATCTCACCGAGCTCCAATTCAAAC
TGTATCAACATGTTCCAAAACCCGCAGTTCAGATTCTCAGGTTATCCGTCAACACACCAT
GCCTCAGCTCTCATGTACGGGACTTGA

>G1782 Amino Acid Sequence (domain in AA coordinates: 166-238)
MQVFQRKEDSSWGNMPTTNSNIQGSSEFSLTKDMIMSTTQLPAMKHSGLQLQNQDSTSS
QSTBEESGGGEVASFGEYKRYGCSIVNNNLSGYIENLGKPIENYTKSITTSSMVSQDSVF
PAPTSGQISWSLQCAETSHFNGFLAPEYASTPTALPHLEMMGLVSSRVPLPHHIQENEP
FVNAKQYHAILRRRKHRAKLEAQNKLIKCRKPYLHESRHLHALKRARGSGGRFLNTKKLQ
ESSNSLCSQMANGQNFMSPHGGGSGIGSSSISPSNSNCINMFQNPQFRFSGYPSTHH
ASALMSGT*

>G184 (327..1937)

TGAATTCTAGCCTTTTTGTAGGCGAATCATCTGGACCGTAAGAGACTCTCTCATCGATA
ATAACCACATAATTTAATCAAACCTTTTCTCTCTTTCTAAGATCTTTTGCTTTGCTCT
TTTCTTTTGTATCTTCTATATATGGAGAAGCACAAAACGGTACTTACTATACGATAC
TGTACGGATCCATCAAACCTGGATTAATTATCAAACCGTACATTTTTATCTTACCTGGCAA
GTTACATTCTAGGGTTTTGGAGAATCCAATCAACAACAAAGAAAATAATCATCGTTACA
ATAATCAGTATCACGCACAGACTTAGATGTTCCGGTTTTCCAGTGAGTCTAGGCGGTTTAC
GTGACGAAGACCGTCACGATCAGATCACACCGTTGGATGACCATCGTGTGGTGGTTGATG
AGGTTGACTTCTTCTCAGAGAAGAGAGATAGGGTTTTACGTTGAGAATCAACGACGACG
ACGACGAAGGCAATAAGGTTCTCATCAAATGGAGGGTTTACGAGTTGAAGAAAACGATC
GTTCCAGAGATGTCAATATCGGTCTGAATCTTCTGACCGCAATACGGGAAGCGATGAGT
CAACGGTGGATGATGGACTATCAATGGATATGGAAGATAACCGTGCAAAGATTGAGAACG
CACAACCTACAAGAAGAGCTCAAGAAGATGAAAATAGAGAATCAAAGGCTAAGAGATATGT
TGAGCCAAGCGACGACCAACTTCAATGCCTTACAAATGCAACTTGTGGCCGTATGAGGC
AACAAGAACAACGTAACCTTTCACAAGATCATCTCTGGAGAGCAAAGCAGAAGGAAGGA
AACGGCAGGAAGTGCATAATCATGGTGCCAAGGCAGTTTATGGACCTTGGGCCGTCGTCTG
GAGCAGCAGAGCATGGAGCCGAAGTGTCTATCTGAAGAGAGGACAACGGTTCGTTACAGTT
CTCTCTCTCGCTTCTAGAAAGTTCCAATCCCCGAGAGAACGGAAGAGGTTGCTTGGAA
GAGAAGAAAGCTCAGAGGAATCAGAGTCTAACGCCCTGGGGAAACCTTAACAAAGTCCCCA
AACATAATCCATCCTCTAGCAATAGCAATGGAACAGAAACGGAAATGTTATTGATCAGT
CGGCCGAGAAAGCCACCATGCGGAAAGCCCGTGTCTCAGTTCTGTCGCCGATCTGAAGCTG
CCATGATAAGCGATGGATGTCAATGGAGAAAGTACGGACAAAAAATGGCTAAAGGAAACC
CGTGTCGCGGGCTTATTATCGTTGCACAAATGGCCGGTGGATGTCAGTTTCGCAAGCAAG
TGCAGCGTTGCGCAGAAAGACAGATCTATTCTCATAACCACCTACGAAGGAAACCACAACC
ATCCACTCCACCAGCCGCTACGGCCATGGCCTCAACAACCACCGCAGCTGCAAGCATGC
TCCTCTCGGGCTCAATGTCGAGTCAAGACGGTTTAAATGAACCCAAACAACTCTTAGCTC
GAGCTATCTTGCCTTGCTCTCAAGCATGGCTACAATCTCAGCCTCCGCACCATTCCCAA
CCATCACATTGGACCTCACCAATTCACCCAACGGTAACAACCTAATATGACCACTAATA
ACCCGTTGATGCAGTTTCGCTCAACGGCCCGGTTTCAACCCGGCAGTTTTCCTCAAGTGG
TTGGTCAAGCTATGTACAAATAACCAACACAGTCCAAGTTTTCTGGTTTACAGTTACCGG
CTCAGCCACTGCAGATCGCGGCCACTTCTCGGTGGCCGAGAGCGTTAGTGCTGCCAGTG
CAGCAATTGCGTCCSATCCAACTTTGCGGCGGCTCTAGCGGCAGCGATCACGTCCATTA
TGAACGGTTCCAGTCATCAAATAATAACCAATAATAATAATGTGGCTACGAGCAACA
ATGACAGTAGGCAATAAGAGTTTTTCAATTTGATGGTTCGATTTTTTTTTTTTGGGG

>G184 Amino Acid Sequence (domain in AA coordinates: 295-352)
MFRFPVSLGGSRDEDHRDQITPLDDHRVVVDEVDFSEKRDVRSRENINDDDDEGNKVL
KMEGRVSEENDRSDVNI GLNLLTANTGSDESTVDDGLSMDMEDKRAKIENALQLEELKK
MKIENQRLRDMLSQATTNFNALQMLVAVMRQEQNRNSSQDHLLESKAEGRKQELQIMV
PRQFMDLGPSSGAAEHGAEVSEERTTVRSGPSPLLESSNPRENKRLLGREESSESE
SNAWGNPNKVPKHNPSNSSNNGNRNNGNVIDQSAABATMRKARVSVRARSEAAMISDGCQW
RKYGQKMAKGNPCPRAYYRCMAGGCPVRKQVQRCAEDRSILITTYEGNHNHPLPPAATA

MASTTTAAASMLLSGSMSSQDGLMNPNTNLLARAILPCSSSMATISASAPFPPTITLDTNS
PNGNNPNMTTNNPLMQFAQRPGFNPAVLQVVGQAMYNQKFSGLQLPAQPLQIAAT
SSVAESVSAASAAIASDPNFAAALAAITSIMNGSSHQNNNTNNNNVATSNNDSRQ*
>G1845 (111..989)
AAGACATAATTTTCTCTGTTTTCTAGCTCTCTCTCTCAAATCTTCCATTGCTCTCTG
TTTTGGCAAATCGTGAAGTCCACGTCTTTAAGGCATCAGTGAAGCAAAGATGGACTTTG
ACGAGGAGCTAAATCTTTGTATTACGAAAGGTAAAAATGTTGATCATTCTTTTGGAGGAG
AAGCTTCTTCCACGTCCCAAGATCTATGAAGAAAAATGAAGAGTCTAGTCGTCCTAAAC
CCTATTTCCAATCCTCTCTCTCTCTTATTCTGTTAGAGGCTTTCCCTTTTTCTCTCGATC
CAACACTTCAGAATCAGCAACAACAACCTCGGATCATACGTTCCGGTACTTGAGCAACGAC
AAGACCCGACAATGCAAGGCCAGAAGCAAATGATCTCTTTAGTCCTCAACAACAACAAC
AGCAGCAGCAGTATATGGCCAGTACTGGAGTGACACATTGAATCTGAGTCCAAGAGGAA
GAATGATGATGATGATGAGCCAAGAAGCTGTTCAACCTTACATCGCAACGAAGCTGTACA
GAGGAGTGAGACAACGCTCAATGGGGAAAAATGGGTGCGAGATCCGTAAGCCACGAAGCA
GGGCACGTCTTTGGCTTGGTACCTTTGATACAGCTGAAGAAGCTGCCATGGCCTACGACC
GCCAAGCCTTCAAATTACGAGGCCACAGCGCAACACTGAATTTCCCGGAGCATTTTGTGA
ATAAGGAAAGCGAGCTGCATGATTCAAACTCGTCGGATCAGAAAGAACTGAAACGCCAC
AGCCAAGCGAGGTTAACTTGGAGAGCAAGGAACCTACCGGTGATTGATGTTGGGAGAGAGG
AAGGTATGGCTGAGGCATGGTACAATGCCATTACATCGGGATGGGGTCTGAAAGTCTCTC
TTTGGGATGATTGGATAGTTCTCATCAGTTTTTCATCAGAAAGCTCATCTTCTCTCTCTC
TCTCTTGTCTTATGAGGCCTTTCTTTTGAAAAAGTTTATAAACCACATTGTGTTGTAGG
TTATAGTTTAGGGTTATGCTCATTGGCATTGGATGGAGGCAATTTTGTGATCTCCCAT
TCCACCACATATCAGTCATTATATGTGTCTACCTTTTCTCTGTATTTCTATCATTATCAT
TGTTTTTATTATGTCTGTATGTGTTTCCCTATTGCTACATACATAGATGTCCTCTTTG
TTCAAAAA
>G1845 Amino Acid Sequence (domain in AA coordinates: 140-207)
MDFDEELNLCITKGNVDHVSFGGEASSTSPRSMKMKSPSRPKPYFQSSSSPYSLEAFPF
SLDPTLQNOQQQLGSYVPVLEQRQDPTMQGQKQMSISFSPQQQQQQQYMAQYWSDTLNL
PRGRMMMSQEAQVQPIATKLYRGVRQRQWGWVABIRKPRSRARLWLTFFDTAEBAAM
AYDRQAFKLGRHSATLNFPEHFVNKESELHDSNSSDQKEPETPQPSEVNLESKELPVIDV
GREGMAEAWYNATTSWGPESPLWDDLDSSHQFSSSSSSPLSCPMRPF*
>G1879 (3..917)
AAATGCCCTTAGAGGCTGTCTATACCCGCAAGATCCATTCCGATATCTCTCCAATTGCA
AAGATTTTATGTTCCACGACTTATACTCTCAAGAAGAGTTTCGTAGCTCAAGATACGAAGA
ACAACATTGATAAGTTAGGCATGAACAGAGCTTTGTGGAACAAGGTAAGGAGGACGATC
ATCAATGGCGAGACTATCATCAGTATCCTTTGTTGATCCCTTCGTTGGGAGAAGAGCTTG
GTCTTACCGCCATTGATGTGAGAGTCACTCTCTCCACAGCACCGGAGGAAGAGGAGGA
GAACGAGAACTGCAAGAACAAAGGAAGAGATCGAGAACCAGAGAATGACTCACATCGCCG
TCGAGAGAAATCGCCGAAACAGATGAACGAGTATCTGGCTGTGCTCCGTTCTCTAATGC
CGTCGTCTGATGCTCAAAGAGGAGATCAAGCGTCGATAGTAGGAGGAGCTATAAACTACG
TGAAGGAGTTAGAGCATATTTTACAATCTATGGAGCGAAGAGAACTAGGACTCATGATC
CCAAAGGAGACAAGACTAGCACTAGCTCGTTAGTGGGTCCATTACAGATTTTTTTCAGCT
TCCCACAATATTCTACAAAGTCATCATCAGATGTACCGGAAAGCTCATCTTACCGGCGG
AGATAGAGGTTACGGTGGCAGAAAGCCATGCGAACATCAAGATAATGACGAAGAAGAAAC
CGAGGCAGCTTCTTAAGCTCATAACTTCTTTACAAAGCCTAAGGCTCACTCTTCTTCATC
TCAATGTCACTCTCCCAACTCCATTCTCTACTCCATCAGCGTCAGGGTTGAAGAAG
GAAGCCAACCTGAATACCGTGGACGACATTGCAACAGCTTTGAATCAAACATAAGGAGGA
TTCAAGAAGAGACATAATTGAGCAAAATAGATTATAATTAACCTGTTTATTTTATTTTA
TTTTGAAATAACTGAAATCAGTTTTCTAATTTTTTTTTTTTTTCACTATTCTCTAATCC
TCCCTATGTAGTTGATTTTTGTCTCTTGAATGAATCAATGGTCATAAAGATCTGAAC
AAAAAATTGAATAAAGAAAATGGTT
>G1879 Amino Acid Sequence (domain in AA coordinates: 107-176)
MPLEAVVYPQDPFGYLSNCKDFMFHDLYSQEEFVAQDTKNNIDKLGHEQSFEQGGKEDDH
QWRDYHQYPLILPSLGBELGLTAIDVESHPPQHRKRRTNRNCKNKEIENQRMTHIAV
ERNRRKQMNELAVLRLSPSSYAQRGDQASIVGGAINVVKLEHILQSMPEKRTTRTHDP
KGDKTSTSLVGPPTDFSFQYSTKSSSDVPSSSSPAEIEVTVAESHANIKIMTKKPK
RQLLKLITSLQSLRLTLHLNVTTLHNSILYSISVRVEEGSQLNTVDDIATALNQTIRRI

QEET*

>G1888 (1..729)

ATGAAGATTGGTGTGCTGTTTGTGATAAAGAAGAAGCTTCGGTGTGTTTGTGTCGGAT
GAAGCAGCTCTTTGTAATGGTTGCGATCGCCATGTTCAATTCGCCAATAAACTAGCCGGG
AAACATCTCCGGTTCTCTCTCACTTCTCCTACTTTCAAAGATGCTCCTCTTTGTGATATT
TGCGGGGAGAGGCGTGCATTATTATTTTCCAAGAAGACAGAGCAATACTATGCAGAGAA
TGTGACATTCCAATACATCAAGCTAATGAGCACACTAAGAAACACAATAGATTCCTCCTT
ACCGGCGTTAAGATCTCTGCTCCCGTCAGCCTACCCAAGAGCCTCCAATTCCTCACTCT
GCTGCTGCATTGGTTCGAGCCAAAACCCGACCAAAATCAGTATCGAGCGAGGTCCCAGC
TCGGCCTCCAATGAGGTATTTACGAGCTCTTCTTCGACGACCACGAGCAATTGCTATTAT
GGGATAGAAGAAAACCTACCATCACGTGAGCGATTTCGGGGTCGGGATCGGGTTGTACAGGT
AGTATATCCGATTATTTGATGGAGACATTACCGGGTTGGAGAGTGGAGGATTGCTTGAA
CACCCTTCTTGTGTCTCCTATGAGGATAACATTATTACTAATAACAATAACAGTGAGTCT
TATAGGGTTTATGATGGTTCTTCACAATTCATCATCAAGGGTTTGGGATCACAACCC
TTCTCTTGA

>G1888 Amino Acid Sequence (domain in aa coordinates: 5-50)

MKIWCAVCDKEEASVFCCADEAALCNGCDRHVHFANKLAGKHLRFSLSPTFKDAPLCDI
CGERRALLFCQEDRAILCRECDIPIHQANEHTKKHNRFLTGVKISASPSAYPRASNSNS
AAAFGRAKTRPKSVSSEVPSSASNEVFTSSSTTTSNCYGIEENYHHVSDSGSGSGCTG
SISEYLMETLPGWRVEDLLEHPSVCVSYEDNIITNNNNSES YRVYDGS SQFHHQGFWDHKP
FS*

>G189 (34..987)

CCACAACCTCTCTCCTTGTAGAGAGAGAGATTTTATGGCGGTGGAGCTCATGACTCGGAAT
TACATCTCCGGCGTCGGAGCTGATAGCTTCGCGGTTCAAGAAGCAGCTGCTTCAGGACTC
AAAAGTATCGAAAATTTATCGGTTTAAATGTCTCGTGATAGCTTAACTCTGATCAGCCA
TCTTCTTCTTCCGCCCTCCGCCCTCCGCCCTCCGCCCGCGCAGATCTGAATCAGCTCGTAAC
ACAACGGCGGACGCGGCTGTTTCAAAGTTTAAAAGAGTCATATCTCTCTTAGATCGAACT
CGAACCGGACACGCCCCGTTTAGACGTGCTCCGGTTCATGTTATTTCTCCGGTCTTTTAA
CAAGAAGAACCAGAACGACGCCGTTTCAGTCTCCTCTCTCCTCCTCCGCCGCAATGATC
CGAAAAGGTTTCGTTTCTTCATCGATGAAAACGATTGATTTCTCATCTCTCTCTCTGTA
ACAACCGAATCAGACAACCAGAAGAAGATTATCATCATCAACGTCCTCTGAAACGGCG
CCGTTTTCGCTCTCAAACTCAAAGCCTCTCCACGACGGTCTCGTCTTCTCAAATCAACA
AAGAGAAAATGTAACCTCGAATCTTCTCACCGGAAAATGCGCTTCCGCTTCTCTCTCC
GGTCGTTGTGTCATGCTCGAAGAAAAGAAAGATAAAACAGAGGAGAATAATTAGGGTCCG
GCGATAAGTGCAAAAATGTCCGATGTACCACCGACGATTATTCATGGAGGAAATACGGA
CAAAAACCAATTAAGGATCTCCACATCCAAGAGGATATTATAAGTGTAGTAGCGTAAGA
GGTTGTCCAGCACGTAAACATGTTGAGAGAGCAGCTGATGATTCGTCATGTTGATTGTT
ACTTATGAAGGAGATCATAATCATTCTCTCTCCGCCGCTGATCTCGCCGAGCCGCCGTT
GCTGATCTTATTTGGAATCGTCTTGAAGAACAATCTTTATTTAAGGCTTTTATAAT
ATAAATTTAGATCCTTACTTAGTGAAGTACTCAAATATGAATGAAATCAATGTAATCAA
AATCAAAAAGCTTTTGTCTAAAAA

>G189 Amino Acid Sequence (domain in AA coordinates: 240-297)

MAVELMTRNYISGVGADSFVQEAASGLKSIENFIGLMSRDSFNSDQPSSSASASASA
AADLESARNTTADA AVSKFRVISLLDRTRTG HARFRRAPVHVISPVLQEEPKTTPFQS
PLPPPPQMIRKGSFSSSMKTIDFSSLSSVTTESDNQKKIHHHQRPSETAPFASQTQSLST
TVSSFSKSTRKCNSENLLTGKASASSSGRCHCSKKRKIKQRRRIIRVPAISAKMSDVPP
DDYSWRKYGQKPIKGSPPHGRGYKCSSVRGCPARKHVERAADSSMLIVTYEGDHNHSL
AADLAGAAVADLILESS*

>G1939 (92..844)

AATCATTAGCTTCTTCTCTCTCTCTCTCACAGAGAGAGTAATCACAAGCCAAAGTGAGA
AAAAGAAAACACTAAACCCAGATCGAAAACCATGCTTATTAACAACAACAACAACAA
CAACAATAACAACGATGGTCTTATGATCTCATCAAACGGAGCTTTAATCGAACAACAAC
ATCAGTCGTTGTGAAGAAACACCGGCGAAAGATCGACATAGCAAAGTCGATGGAAGAGG
GAGAAGAATCCGTATGCCGATTATATGTGCTGCTCGTGTGTTTTCAGCTAACGAGAGGCT
TGGTCATAAGTCAGATGGCCAAACAATTGAATGGTTACTTCGTCAAGCAGAGCCTTCTAT
TATAGCTGCAACAGGAACGGTACAACCTCCAGCGAGTTTCTCAACTGCTTCTGTCTCTAT
CCGTGGAGCCACCAATTCTACTTCTTTAGATCATAAACCCACTTCTTTACTTGGTGGTAC

GTCACCGTTTATACTTGGGAAACGTGT'TAGAGCTGATGAGGATAGTAATAATAGTCATAA
TCATAGTTCTGTTGGTAAAGATGAGACCTTTACGACAACACCAGCTGGGTTTTGGGCTGT
TCCGGCGAGGCCGGATTTTGGACAAGTTTGGAGTTTGTCTGGAGCTCCAACAAGAGATGTT
TTTACAACAACAACATCATCATCAGCAACCATTGTTTGTTCATCAGCAACAGCAACAACA
AGCTGCAATGGGTGAAGCTTCTGCTGCTAGAGTTGGGAATTATCTTCCGGGTCATCTTAA
TTTGCTTGTCTCTTTATCCGGTGGATCTCCCGGGTCGGATCGAAGAGAGGAAGATCCACG
TTAATGGTTTAAAGCCCTTTTAGGTTTGAGGGCAAAATTTGGTATATATATTTATTATCTT
CTCTTCTCTATTGTTGTCTATTGTTTCTCTATGTGTGTGTTTTAGTGTGTTAGAGATTGA
TTTGGTTTTCAGAATCTCTGCAAGTGATTGAGAGTTTTCGTTAGCTTTAAGTAAGTTAAA
GACGGTTGTTTTGATTAGGGTTAAATTAGGGTTTAAAGAATCTGTTGTTTTTTTGGAGGG
AGATCGATTCTTATCGGATCCAAGATTACTTTTAGGAAAAAAGGAAAAATTCAGAAAC
CACGGTGGTTCTTTTCTCTTTTTTTTTTTT

>G1939 Amino Acid Sequence (domain in AA coordinates: 40-102)
MSINNNNNNNNNNDGLMISSNGALIEQQPSVVVKPPAKDRHSKVDGRGRRIRMPIICA
ARVFQLTRELGHKSDGQTIEWLLRQAEPSIIAATGTGTPASFSTASVSIRGATNSTSLD
HKPTSLGGTSPFILKRVRAEDSNNSHNHSSVGKDETF'TTPAGFWAVPARPDFGQVW
SFAGAPQEMFLQQQHHHQQPLFVHQQQQQQAAMGEASAARVGNYPGHLNLLASLSGGSP
GSDRREEDPR*

>G194 (192..1205)
TCTTTCTTCTCTCTCTCTCTCTCTTGAACCCCTAAAAACTCTTTCTTTACAAGGATT
GATCTTTTTGTATTTTTGATTTTTGACATTTGCTTTGTGTTTCGATCTCTGTTTTGATGCGA
TTTCTCTGTTTTTAAAGCCATTGATAGATTGTTCCGGTAAAGCTCAGCGAGAGAAGAA
GAAGAACAACATGGAGTTTACAGATTTCTCAAAGACGAGTTTTTACTACCCGTCGTCAC
AAAGCGTTTGGGATTTTCGGAGATTTAGCGGCGCGGAGAGGCATTCTTTAGGGTTCATGG
AGTTATTAAGTTCTCAGCAGCATCAAGACTTTGCTACTGTTTCTCCTCATTCCTTCCTTC
TCCAACGCTCTCAACCGCAAACGCAACGCAACCATCGGCGAAGCTGTCTTCAAGTATCA
TTCAAGCTCCACCGTCAGAGCAATTAGTGACGTCAAAGGTGGAGTCTTTGTGTTCCGATC
ATTTGTTGATAAACCACCGCGACTCCTAACTCGTCATCGATTTTCGTCTGCTTCAAGCG
AGGCTCTAAATGAAGAGAAACCGAAAAACAGAGACAATGAAGAAGAAGGAGGTGAAGATC
AACAAGAGAAGAGTCATACTAAGAAACAGTTGAAAGCAAAGAAGAATAATCAGAAGAGAC
AGAGAGAGGCAAGAGTCGCATTATGACAAAGAGTGAAGTTGATCATCTCGAAGATGGTT
ATCGCTGGCGAAAAATATGGTCAAAAAGCTGTCAAAAACAGTCTTTTCCAGGAGTTACT
ACCGTTGCACAACGGCTTCATGTAACGTGAAGAAGAGAGTGGAGAGATCATTCAGAGATC
CAAGCACTGTGGTTACAACCTACGAAGGTCAACACACTCACATTAGTCCACTCACGTCTC
GTCCTATTTCCACTGGAGGTTTCTTCGGATCGTCAGGAGCTGCTTCGAGTCTCGGTAATG
GTTGCTTTGGGTTTCTATTGATGGCTCCACGTTAATCTCTCCTCAGTTCCAACAGCTTG
TCCAATACCATCACCACAGCAGCAACAAGAACTCATGTCTTGTGTTTGGAGGAGTCAACG
AGTACCTTAATAGCCACGCTAATGAGTATGGTGATGATAATCGTGTGAAGAAGAGTCGAG
TTTTGGTTAAAGATAATGGACTTCTGCAAGATGTTGTTCCGTCTCATATGTTGAAGGAAG
AGTAGTAGTATATATATAGTCTTATAGTTTAAATCTAGTTTTTTTTTGTATAATTGTCTA
AAAGAAACGGATCTTTGTTCTGATGAAGAAGATGTTTCTTATGGTTCTGAAATCGTAA
GGTAATGATGATTGTACCAAGCCGAGAAAGTACTTGTGATTTTCACCATGAATCACTAT
AAATGTAATTTTTATTACTGTGAAAAAAAAAAAAAAAA

>G194 Amino Acid Sequence (domain in AA coordinates: 174-230)
MEFTDFSKTSFYYPSSQSVWDFGDLAAERHSLGFMELLSSQQHQDFATVSPHSFLLQTS
EPQTQTQPSAKLSSSIQAPPSEQLVTSKVESLCSHDLLINPPATPNSSSISSASSEALN
EEKPKTEDNEEEGGEDQKEKSHTKQLKAKKNNQKRQREARVAFMTKSEVDHLEDGYRWR
KYGQKAVKNSPFPRSYYRCTTASCNVKKRVERSFRDPSTVVTTYEGQHTHISPLTSRPI
TGGFFGSSGAASSLGNCGFGFPIDGSTLISPQFQQLVQYHHQQQQQELMSCFGGVNEYLN
SHANEYGDDNRVKSRVLVDNGLLQDVVPSHMLKEE*

>G1943 (137..1858)
ACATTTGTTTCTAATCTCAGACATAAATAATTTTTGTTCCCGACTTCAAAACCAACGATG
ATTATATCATTCACATTCATTTCTTCTACTTCTTCCTTCTCCTTGATCTCATTTCCCT
AGAAAATCCATCTATCATGGGTGAAGATGATATAGTGGAGCTCTTATGGAAGAGTGGCCA
AGTCGTTAGAACCAGTCAAACACAGAGACCTCCTCCAATACACCACCATCTTCTCTCC
ACCACCCATTCTTCGTGGTAGCGGAAGCGGCAACGGAGAAGAAAATGCCCCGCTTCCACT
TCCACAGCCTTCACCTCCCTCCATCATCAGAATCTTTTCATTCTGGAAGACGAAATGTC

TTCTTGGCTTCACCATCTCACCCCGGCGTTACGTCCACCCCGGCTTCTTCTGTCTCCCT
GCCACCACCACCAATGCTCCGCGTGAAGATGATATAGTGGAGCTTTTATGGCAAAGCGG
CCAAGTAGTTGGAACCAACCAACACATAGACAATCCTACGATCCTCCTCCATTCTCCG
CGGCAGCGGAAGTGGCAGAGGAGAAGAAATGCTCCCTTTACAACTCCGCTCACCT
GCATCAGCAAAATCTCTTCAATCAAGAAGGCGAAATGTATTCTGGCTACACCATCTTA
CCGCCAAAATATTTCTGCTCAGAACTTCTCAACTCCACTCCGGCTACTCACCCGCAAAG
TTCCATCTCTCTGGCACCACGTCAGACTATCGCCACGAGAAGGGCGGAAAACCTTTATGAA
CTTCTCGTGGCTAAGAGGGAACATATTTACCGCGGTAGAGTTGATGAAGCTGGACCGTC
GTTTTCTGGTGAAGAGAATCGATGCAGGTAGGCTCGAACACGACCCCCCTTCTTCTC
TGCCACTGAATCATGTGTAATACCAGCTACAGAGGGCACCGGAGTCGAGTGTGGGAAC
TTTGGCAGCTCATGATCTTGGTCGGAAGGGAAAGGCGGTGGCGTTGAGGCGCCGGAAC
ACCATCTTCAGGAGTGTGCAAGGCCGAAACAGAGCCGGTTCAGATACAACAGCAACGGA
GTCAAGCTAAAAGCGAGAGAAGAAACCATGGAAGTGAAGAAGCTCGTGGTTCAACGTC
TAGAAAGAGATCAGCAACTGCAGAAATGCATAACCTCGCCGAAAGGAGAAGGAGAGAAA
GATCAACGAGAAGATGAAGACTCTGCAACAACCTATTCTCGCTGCAACAAGGTTGAATC
TGATTCTGTTTCTACTCTGATCAGTCTACTAAAGTTTCAACCGTGGATGATGCTATCGAG
TACGTCAAATCGTTACAGAGCCAAATACAAGTATGCTCTTCAAAACAGAATGTGTTTTAA
ACCAATGGTTCAACATGGAAGAGTTTATATGTATCTAGTTTGTGTTGAGATGATGTCGAC
GGGACAGGTATGATGTGCGCAATGATGAATGCGGGAATACGCAACAGTTTATGCCCCA
TATGGCCATGGATATGAACCGACCTCCTCCATTATACCTTTCCCGGCACATCTTTTCC
TATGCGGCTCAAATGGCAGGTGTAGGTCCATCATATCCAGCACCGCGTACCCTTTTCC
CAACATTAGACCTTTGACCCATCCAGAGTCCGTTTACCAAGCCCGCAGCTAACCCGGT
GTCGAACAGCTCAGTTTCCGGCTTACATGAATCCCTATAGCCAGTTTGTGGTCCCCA
CCAGTTGCAACAACCTCCTCCTCCTCATTTCAGGGTCAAACAACATCACAACCTGAGTTC
CGGGCAGGCAAGTAGTAGCAAGGAACCTGAGGATCAGGAGAACCAACCAACAGCTTAGTT
AAAGTGTGGAGCTGAAACCGATCAGTTCTTCAAGCAAATTACAACCTTTGAAGATAAACCA
GAGTTGTAACATGTAGATTTTGTCTGTTAAGTTTAAATGTAAGTACTTTTGTAGTTAATGGG
AAAGATACTGACAGGTGCAAGGTGGTCAATTTGTGCATCACGCTTAAGATTCCTCGA
TGTGGCCAGTATCTCCCTTTCTAGCATGTGAGGTCCCTACTCTCTGTTTCTACGGAGAC
CAAATGTTCCAGCTGATTAAACACACAATGACTTACCAAAAGTACACGCGGCCCATCCTCG
TCTTTATGTTCCAAGTGCAGCTGTTTGTATTGTAAGCATTTTTCTTATAATAATAAA
ACAGCTCTATCTCGTTAAAAAAA

>G1943 Amino Acid Sequence (domain in AA coordinates: 335-406)

MGEDDIVEILLWKSQVVRTSQTRPSSNTPPSLPPPPILRGSGSGNGEENAPLPLPQSP
PLHHQNLFILEDEMSSWLHSHPGVSTSTPASSVSLPPPPNAPREDDIVELLWQSGQVVG
NQTHRQSYDPPPIILRGSGSGRGEENAPLSQPPPHLQONLFIQEGEMYSWLHHSYRQNYF
CSELLNSTPATHQSSISLAPRQTIATRRRAENFMNFSWLRGNIFTGGRVDEAGPSFSVVR
ESMQVGSNTTPPSSSATESCVIPATEGTASRVSGTLAAHDLGRKGKAVAVEAAGTPSSGV
CKAETEPVQIQPATESKLKAREETHGTEEARGSTSRKRSRTAEMHNLAERRRREKINEKM
KTLQQLIPRCNKVESDSVSTLISLLKFQRMMLSSSTSNRYRAKYKYLQNRMCFKPMVQH
GKSSYVSSFVEMMSTGQGMSPMMNAGNTQQFMPHMAMDMNRPPPIPFPGTSFMPAQM
AGVGPSYPAPRYPFPIQTFDPSRVRLPSPQPNVSNQPFQPAYMNPYSQFAGPHQLQQP
PPPPFQQTTSQLSSGQASSKEPEDQENQPTA*

>G21 (79..966)

TGTGGAGGAATATTAATACAGCCCACTTCACATCTATTTTGTGCAACCATCTCTCTAAA
GCTTCTTCTCTCATAACAATGGCAAGACAAATCAACATAGAGAGTAGTGTCTCAAGTT
ACCTTTATCTCCTCCGCCATCCCCCGGTATCTTCTCCTCCTCCATCACCGCTCCCGCC
TCATTGTCTCTTGACCTACTACATCTTCTCTTCTTCTCGTCATCAACAAATTCTAACTTC
ATTGAGGAAGACAACCTCTAAAAGAAAAGCATCTCGAAGATCATTGTATCGTTAGTCTCC
GTTGAAGACGATGATGATCAAAACGGTGGAGGTGGGAAACGGCGAAAGACCAACGGTGG
GATAAACATCCGACGTATAGAGGAGTGAGGATGAGGAGTTGGGGAATGGGTGTCTGGAG
ATTAGAGAGCCGAGAAAGAAATCAAGAATCTGGCTCGGGAATTATCCAACGGCTGAGATG
GCAGCTCGAGCTCATGACGTAGCGGCTTTAGCCATTAAAGGTACAACGGCTTACCTCAAT
TTTCCCAAGTTAGCCGCGGAGCTTCTCGTCCAGTCACAAATTCTCTAAAGACATTCAA
GCCGCCCGCTTTTAGCGGCGTTAACTGGCAAGATTCGGTCAACGATGTGAGTAATTCT
GAAGTGGCTGAAATAGTTGAAGCCGAGCCGAGTCGAGCCGTGGTGGCTCAGTTGTTTCT
TCGGACACAAGCACGACGACGACTCAGAGTCAAGAGTATTCCGAAGCTTCTGTGTCT

TCGACTTCGGCGTGCTACGGACAAAGACAGTGAGGAAGAGAAGCTGTTTGATTGCGCGGAT
TTGTTTACCGATGAGAATGAGATGATGATACGAAACGATGCGTTTTGCTACTACTCGTCC
ACGTGGCAGCTTTGTGGAGCCGATGCTGGGTTTCGGCTTGAAGAGCCGTTTTTCTATCT
GAATGACTAAAGTACCCCTCTCGAGAGAGCTCTCACTAACACT

>G21 Amino Acid Sequence (domain in AA coordinates: 97-164)
MARQINIESSVSQVTFISSAIPAVSSSSSITASASLSSSPTSSSSSSSTNSNFIEEDNS
KPKASRRSLSSLVSVEDDDDDQNGGGGKRRKTNGGDKHPTYRGVVRMSWGKVVSEIREPRK
KSRIWLGTPTAEMAARAHDAALAIKGTAYLNFPLAGELPRPVTNSPKDIQAAASLA
AVNWQDSVNDVSNSEVAEIVEAEPSRAVVAQLFSSDTSTTTTTQSQEYSEASCSTACT
DKDSEEEKLFDLPDLFTDENEMIRNDAPCYSSSTWQLCGADAGFRLEEPFFLSE*

>G2132 (42..1031)

ATTCTGTACTTAGTACCGGAGTTTAGTCCGAGAGAGACAATGATCAGTTTCAGAGAAG
AGAATATCGATCTCACTTGATTAACAATAGTGTAATCTGTAATGATCCAGACGCCA
CCGATTCTCTAGCGACGATGAATCTATCTCCGGCAATAATCTCGCCGTCAGATCAAAAC
CAAAACCACCGAAACGTTACGTCTCAAAGATCTGTGTCCGACGCTGATCAAAAGGTATG
AGAACGTTTCGAATTCTACAGGGAATAAAGCAGCCGGAACCGGAAACGTCGTCGGGTT
TCAAAGGCGTACGACGAGGCGGTGGGGGAAATTTGCGGCGGAGATAAGAAATCCGTTTG
AGAAGAAGAGAAAGTGGCTTGAACGTTTCTACTGAAGAAGAAGCAGCAGAAGCTTACC
AAAAGAGTAAAGAGAGTTTGTGTAACGATTGGGTTTAGTTAAACAGGAAAAAGACCTAG
TAGATTGACCAAGCCGTGCGGTGTACGTAAACCAGAAGAGAAGGAAGTTACTGAGAAGT
CGAATTGCAAAAAGGTAAATAAGAGAATTGTTACTGATCAGAAGCCATTGGTTGTGGTT
ATAACGCTGATCATGAAGAAGAGGGAGTGATTAGTAAATGTTGGAAGATCCGTTGATGA
CATCGTCAATTGCTGATATTTTGGTGATTCGGCTGTGAAGCAAATGATATTTGGGTGG
ATTACAATTAGTGAATTTATTTCCATTGTAGATGATTTCAAGTTTGATTTTGTGGAGA
ATGATAGAGTAGGAAAGGAGAAAAACATTGGGATTTAAGATTGGGGATCACACTAAAGTTA
ATCAACATGCCAAAATCGTATCGACCAATGGGGACTTATTCGTCGATGATTTACTTGATT
TTGATCCGTTGATAGATGATTTAAGTTAGAAGATTTTCTATGGATGATCTTGGATTAT
TAGGAGATCCAGAGGATGATTTTAGTTGGTTTAAATGGTACTACTGATTGGATCGATA
AGTTTTTATGAATACTTTCTTGACACGGCCAACGGTATTAGTAC

>G2132 Amino Acid Sequence (domain in AA coordinates: TBD)
MISFREENIDLNLIKTISVICNDPDATSSSDDESISGNPNRRQIKPKPKRYVSKICVP
TLIKRYENVSNSTGNKAAGNRKTSSGFKGVRRRPWGKFAAEIRNPFKKRWLGFPTTEE
EAAEAYQKSKREFDERLGLVKQEKDLVDLTKPCGVKRPKEEKEVTEKSNCKKVNKRIVTDQ
KPFGCGYNADHEEEGVISKMLEDPMTSSADI FGDSAVEANDIWDVNSVEFISIVDDF
KFDFVENDRVGKEKTFGFKIGDHTKVNQHAQIVSTNGDLFVDDLLDFDPLIDDFKLEDFP
MDDLGLLGDPEDDDDFSWFNGTTDWIDKFL*

>G2145 (1..777)

ATGGACGTTTTTGTGTGATGGTGAATTGGAGTCTCTCTTGGGGATGTTCAACTTTGATCAA
TGTTTCATCATCTAAAGAGGAGAGACCGCGAGACGAGTTGCTTGGCCTCTCTAGCCTTTAC
AATGGTCATCTTCATCAACATCAACACCATAACAATGTCTTATCTTCTGATCATCATGCT
TTCTTGCTCCCTGATATGTTCCCATTTGGTGCAATGCCGGGAGGAAATCTTCCGGCCATG
CTTGATTCTTGGGATCAAAGTCATCACCTCCAAGAAACGCTTCTCTTAAGAGGAAACTA
CTTGACGTGGAGAATCTATGCAAACTAACTCTAACTGTGACGTCACAAGACAAGAGCTT
GCGAAATCCAAGAAAAACAGAGGGTAAGCTCGGAAAGCAATACAGTTGACGAGAGCAAC
ACTAATTGGGTAGATGGTCAGAGTTTAAGCAACAGTTCAGATGATGAGAAAGCTTCGGTC
ACAAGTGTTAAAGGCAAACTAGAGCCACCAAGGGACAGCCACTGATCCTCAAAGCCTT
TATGCTCGGAAACGAAGAGAGAAGATTAACGAAAGGCTCAAGACACTACAAAACCTTGTG
CCAAACGGGACAAAAGTCGATATAAGCACGATGCTTGAAGAAGCGGTCCATTACGTGAAG
TTCTTGACGCTTCAGATTAAGTTGTTGAGCTCGGATGATCTATGGATGTACGCACCATG
GCTTACAACGGCCTGGACATGGGGTTCATCACAACCTTTTGTCTCGGCTTATGTGA

>G2145 Amino Acid Sequence (domain in AA coordinates: 166-243)
MDVFDGGELESLLGMFNFDQSSSKKEERPRDELLGLSSLYNHLHQHQNHNVLSSDHHA
FLLPDMFFPGAMPGGNLPAMLDSDQSHHLQETSSLRKLLDVENLCKTNSNCDVTRQEL
AKSKKKQRVSSSENVTNVDGQSLSNSDDDEKASVTSVKGKTRATKGTATDPQSL
YARKRREKINERLKTQLNLVPNGTKVDISTMLEAVHYVKFLQLQIKLLSSDDLWMYAPL
AYNGLDMGFHNNLLSRML*

>G23 (22..732)

TATCAAACGAGAGTACAAAAGATGACGCTCACTCAACAGCTCTGCATCACCAACATCATCG
 TCATCAGACCAATCTGATGCACTACTACAACAAGCACCCACTTGTCTGAAGAAGAAGCT
 CCACCCAGAAACAACAACAAGAAAGAGAAGGAGAGATTCTTCTTCTGCTTCTTCATCT
 TCTTCAATGCAACATCCTGTTTACAGAGGTGTGCGGATGAGAAGTTGGGGCAAATGGGTC
 TCCGAGATCCGACAACCTCGTAAGAAAACCTCGTATTTGGCTCGGCACCTTTTGTACCGCT
 GATATGGCTGCTCGTCTCACGACGTCGCTGCTCTCACCATCAAAGGCTCCTCCGCCGCTC
 TTAAATTTCCCTGAGCTTGCTTCTCTCTTCCCTCGTCCGGCGTCATCATCGCCGCATGAT
 ATCCAGACAGCCGCCGAGAAGCCGCCCGCCATGGTGGTGAAGAAAACTGTTAGAGAAG
 GATGAGGCTCCGGAGGCCACCTTCGTGGAATCTTCTTACGTGGCGGCGAGTCAGAG
 GATGAGGAGAGGTTGGAGAAAATTGTGGAGCTGCCAATTAAGGAAGGAAGTTATGAC
 GAGAGTGTGACATCACGTGCTGATCTGGCTTATCTGAGCCGTTGATTGTTGGGTGTAT
 CCTCCGGTTATGGATTTTATGAAGAAATATCGGAGTTAATTCGTGGAATTGTGGAGC
 TTTAATCACTAATTAAGTTAGGAAAGTGCATTATATGCAATATGCATCATAGATAACA
 TTGTATTTCTTTTCTTTTGTACGGATACGTAGCATATGCTACTATACTAGGGCTAGTG
 TACCAATATTTGTAATAATATACTTATTAATATTTATGTAATGTGTAATATATAACAT
 ACAATTATTGTAAGTTTGGAAATTGGAAACTATCGTTACGCAATGTTCTTGTAACAAAAA
 AAAAAA

>G23 Amino Acid Sequence (domain in AA coordinates: 61-117)
 MTSLNSSASPTSSSSDQSDATTTTSTHLSEEEAPPRNNTRKRRRDSASSSSSMQHPV
 YRGVRRMSWGKVVSEIRQPRKKTRIWLGTFTVADMAARAHDAALTIKSSAVLNFPELA
 SLFPRPASSPHDIQTAAEAAMVVEEKLLEKDEAPEAPSSSESSVAAESEDEERLEK
 IVELPNIEEGSYDESVTSRADLAYSEPPDCWVYPPVMDFYEEISEPNFVELWSFNH*

>G2313 (104..724)
 CGTCGACACAATCGCTCTCCGTAACATATTCCACAAAACGATCTTCTTGTTCCTTGAAT
 TTTTAGCCATCTCTTTTTTTTTTTTCTCATTTTCTCGGATACTATGGCTTCGAGTCCACG
 CTGGACGGAGGACGACAACAGGCGTTTTAAGTCAGCTCTGTGCAATTCCCTCCGGATAA
 CAAGCGTTTGGTGAATGTGCGCCAGCATCTGCCGAAACCTTTGGAGGAGGTGAAGTACTA
 CTACGAAAAGTTGGTCAACGATGTTTATCTGCCGAAACCTTTAGAGAATGTCACCCAGCA
 TCTGCAGAAACCTATGGAATGGAGGAGATGAAGTACATGTACGAAAAGATGGCCAACGA
 TGTTAATCAGATGCCCGAGTACGTACCACTGGCGGAATCGAGTCAGTCCAAACGCAGGAA
 GAAGGATACGCCAAATCCTTTGGACAGAAGAGGAACACAGATTGTTTCTGCAAGGATTGAA
 AAAGTATGGGGAAGGAGCTTCGACGTTGACATCAACGAATTTTGTGAAGACAAAGACTCC
 ACGGCAAGTGTCAAGCCATGCACAGTATTACAAAAGGCAAAAATCGGACAATAAGAAGGA
 GAAACGCCGAGTATTTTTGACATAACTTTGGAGTCTACCGAGGGCAATCCAGATTCTGG
 AAATCAGAACCCTCCGGATGATGATGATCCGTCCCAAGGTCAAGGCATTGTCTTGGAGT
 TTAGATGTTGGAAGATAGAAGAATGGTGTGAAAGC

>G2313 Amino Acid Sequence (domain in AA coordinates: TBD)
 MASSPRWETEDNRRFKSALSQPPDNKRLVNVAQHLPKPLEEVKYYYEKLVDVYLPKPL
 ENVTOHLQKPMEMEEMKMYEKMANDVNQMPEYVPLAESSQSKRRKKDTPNPWTEEEHRL
 FLQGLKKYGEASTLTSTNFVKTKTPROVSSHAQYKQKSDNKKERRSIFDITLESTE
 GNPDSGNQNPDDDDPSQGQGTCLGV*

>G2344 (1..573)
 ATGACTTCTTCAATCCATGAGCTTTCTGATAACATTGGAAGTCATGAGAAGCAAGAACAG
 AGAGATTCTCATTTCACCAACCAATCCCTTCTGCAAGAAATTATGAATCAATTGTTACA
 AGTTTAGTCTACTCAGACCCGGGACTACAAATTCATGGCACCTGGACAATATCCATAT
 CCAGATCCTTACTACAGAAGCATATTTGCACCGCCTCCACAACCGTATACCGGGGTACAT
 CTACAGTTGATGGGAGTGCAGCAACAAGGCGTTCTTTTACCATCTGATGCAGTCGAGGAA
 CCTGTTTTTGTAAACGCAAGCAATACCACGGTATACTAAGGCGCAGACAATCAAGAGCA
 AGACTTGAGTCTCAGAATAAAGTCATCAAGTCACGTAAGCCGTATTTGCAATGAATCTCGG
 CATTTGCATGCGATAAGACGACCAAGAGGATGTGGCGGGCGGTTTCTAAATGCCAAGAAG
 GAGGATGAGCATCACGAAGACAGTAGTCATGAAGAAAAATCCAACCTTAGCGCTGGTAA
 TCCGCCATGGCTGCTTCTAGTGGTACATCTTGA

>G2344 Amino Acid Sequence (domain in AA coordinates: TBD)
 MTSSIHESLSDNIGSHEKQEQRDSHFQPPIPSARNYESIVTSLVYSDPGTTNSMAPGQYPY
 PDPYYRSIFAPPPQPYTGVLHQLMGVQQQGVPLPSDAVEEPVFNQKQYHGLRRRQSR
 RLESQNKVIKSRKPYLHESRHLHAIRRPGRGGRFLNAKKEDEHHEDSSHEEKSNLSAGK
 SAMAASSGTS*

GTGATCCAAGCAGTGTAAATCACCACCTTACGAAGGTCAACATACTCATCCTCGTCCACTAC
TCATCATGCCCCAAGAAGGCAGCTCTCCATCCAATGGCTCAGCTTCTAGGGCCACATTG
GCCTCCCTACACTCCCTCCTCAGCTTTTAGATTACAACAACCAACAACAACAAGCGCCGT
CTTCTTTTGGAAACCGAGTACATTAACAGGCAAGAAAAAGGAATTAATCATGATGATGATG
ACGATCATGTTGTGAAGAAGAGTCGAACCTCGGGATCTGCTGGATGGAGCTGGTTTAGTCA
AAGATCATGGCCTTCTTCAGGATGTTGTTCCCTCTCATATCATTAAAGGAAGAGTATTAGT
TAATCGCATAATTATGTAGCTAGCTAGCTAG

>G2517 Amino Acid Sequence (domain in AA coordinates: TBD)

MENVGVGMPFYDLGQTRVYPLLSDFHDLAERYPVGFMDLLGVHRHTPTHTPLMHFPTTP
NSSSSEAVNGDEEEEDGEEQKHKKRKFKFTKMSRKQTKKKVPKVSFTRSEVLHLLDDG
YKWRKYGQKPVKDSFPFRNYRCTTTWCDVKRVERSFSDPSSVITTYEQHHTHPRPLLI
MPKEGSSPSNGSASRAHIGLPTLPPQLLDYNNQQQAPSSFGTEYINRQEKGINHDDDDD
HVVKSRTRDLLDGAGLVKDHGLLDVVPISHIIKEEY*

>G2521 (103..768)

ATTCTCCACAATTTTCATAACTTTCTTCCGCTCAACTTCAGATAAATTCGGATTCTGTAGC
TCTTTCAATACGACTGCGGAGATCAGAGCCAATTATTTGGTTATGGCGTCTCTGATCTCA
GATATTGAACCGCCGACGAGTACTACTTCAGATCTCGTTCGGAGAAAGAAGAGATCCTCT
GCTTCATCCGCCGATCGTCTCGTTCAAGCGCATCTTCCGTCTCCGGTGAGATTACGCG
CGATGGCGATCGGAGAAGCAACAACGGATCTACTCAGCCAACTGTTCCAAGCGCTCCAA
CAAGTCCGCCTCAACTCTTCCGCCTCAACATCATCATCTCCAACGGCTCAGAAACGAGGA
AAGGCCGTCCGTGAAGCCGCCGATCGAGCTCTTGCCGTTTCCGCTCGGGAAGAACA
TGGAGCAGAGCGATCTTAGCTAATCGGATCAAACTGAAATTTTCGTAAACAGAGACGCTCT
CGAGCTACGATGGCGATTCCGGCCATGACTACGGTGGTTAGTAGCAGCAGCAACAGATCG
AGAAAACGGAGAGTGTCCGTTGTGAGATTGAATAAGAAGAGTATACCGGATGTTAACC
AAAGTACGTGTTCTAGGCCGTTAGTTCCCGGTTGCGGTAAACAATCCGTACCGGTGATT
CTAGAAGAAGCAACTGATTATATTACGGCTCTGGAGATGCAAGTGAGAGCCATGAATCT
TTAGTTACGCTTCTCTCTCTACGGCTCAGCTCCTCCACCGATTTGATGAGGTTAAAT
CGTCTTTTTTAATCTACCATCTCTCGATCTTTCACAGCTTATGTGTATATAGAAGATT
GTTTGATTATAATCTGTAATACTCTTCCCAACCGCTGATTCTTCTCTGCTACAAGTAA
AGTAAATTTGAACCGAGTCTTCCCATTTTACGATCCTCAAGTCTAAATTAAGTATATG
ATTGATTAATAAAGTCTTTACCATTAGGGTTC

>G2521 Amino Acid Sequence (domain in AA coordinates: 145-213)

MASLISDIEPTSTTSDLVRRKRSSASSAASSRSSASSVSGEIHARWRSEKQRIYSAK
LFQALQOQVRLNSSASTSSSPTAQKRGKAVREAADRALAVSARGRTLWSRAILANRIKLK
RKQRRPRATMAIPAMTTVVSSSNRSRKRVSVLRLNKKSIIPDVNRKVRVLGRLVPGCGK
QSVFVILEEATDIYIQALEMQVRAMNSLVQLLSYGSAPPI*

>G258 (60..983)

AGTGACCACCTGTGGTTAATCAACACCAAGAGACCTTGTAATATATAAGTTAGGAAGA
TGAGAGAGAAGTGGGAAATGAAAAGAGATGAAATGGGACATCGATGTTGTGGAAAACACA
AAGTGAAGAGAGGTCTTTGGTCTCCAGAGGAAGACGAGAAGCTTCTCGTTATATACCA
CTCATGGTCATCCTAGTTGGAGTTCGTTCCAAAGCTTGCCGGGTTGCAGAGATGTGGGA
AGAGTTGCAGATTAAGGTGGATAAACTATCTAAGGCCTGATCTGAGGAGAGGTTCTGTTA
ATGAGGAAGAAGAGCAGATTATCATCGACGTACATCGTATCTTGGTAACAAATGGGCTC
AGATTGCTAAGCACTTACCTGGACGCACTGATAATGAAGTCAAGAACTTTTGGAACTCAT
GCATTAAGAAGAACTTCTTTCTCAAGGCTTAGATCCTTCTACACATAATCTTATGCCTT
CACACAAAAGATCTTCTTCTCAAAACAATAATAATATCCCCAAGCCAAACAAAACGACGT
CCATCATGAAGAACCCTACTGATCTTGATCAATCAACCACTGCTTTTTCAATCACAACA
TCAATCCACCCACTTCCACTAAACCAAACTTAAATCTCCTAACCAGACTACAATCC
CATCTCAAACCGTGATCCCTATCAATGATAACATGTCAAGTACTCAAACCATGATCCCTA
TCAATGATCCCATGTCAAGTCTTTAGATGATGAGAATATGATTCCCTCACTGGTCAGATG
TTGATGGAATGGCGATCCACGAAGCTCCGATGTTGCCTAGTGATAAGGCAGTAGTGGGAG
TGGATGATGATGATCTCAACATGGACATTTTGTAAACTCCTTCTTCTCTGCTTTTG
ATCCTGATTTTGCTTCCATTTTCTCCTCTGCAATGTCTATCGATTTCAATCCCATGGATG
ATCTTGGCAGCTGGACCTTTTAGCTTTTACTCTACAGC

>G258 Amino Acid Sequence (domain in AA coordinates: 24-124)

MREKWEKMRDEMGRCCGKHVKRGLWSPEEDEKLLRYITTHGHPSSVPLAGLQRCG
KSCRLRWINYLRPDLRRGSFNEEEQIIIDVHRILGNKWAQIAKHLPGRTDNEVKNFWNS

CIKKLLSQGLDPSTHNLMPSHKRSSSSNNNNIPKPNKTTSIMKNPTDLQSTTAFSITN
INPPTSTKPNKLKSPNQTTIPSTQVIPINDNMSSTQTMIPINDPMSSLLDDENMIPHWSD
VDGMAIHEAPMLPSDKAVVGVDLNDILFNTPPSSAFDPDFASIFSSAMSIDFNPM
DLGSWTF*

>G280 (108..722)

AAGTTAATATGAGAATAATGAGAAAACCACTTTCCCAAATTGCTTTTTTAAATCCCTCCT
CACACAGATTCCCTTCCTTCATCACCTCACACACTCTCTACGCTTGACATGGCCTTCGATC
TCCACCATGGCTCAGCTTCAGATACGCATTTCATCAGAAGTCCGCTCGTTTTCTCTCCAC
CTTATCTCAGATGATAATGGAAGCGATTGAGTCCTTGAACGATAAGAACGGCTGCAACA
AAACGACGATTGCTAAGCACATCGAGTCGACTCAACAACTCTACCGCCGTCACACATGA
CGCTGCTCAGCTACCATCTCAACCAGATGAAGAAAACCGGTCAGCTAATCATGGTGAAGA
ACAATTATATGAAACAGATCCAGATGCTCCTCCTAAGCGTGGTCGTGGCCGTCCTCCGA
AGCAGAAGACTCAGGCCGAATCTGACGCCGCTGCTGCTGCTGTTGCTGCCACCGTCTG
TCTCTACAGATCCGCTTAGATCTCGTGGCCGTCACCGAAGCCGAAAGATCCATCGGAGC
CTCCCCAGGAGAAGGTCATTACCGGATCTGGAAGGCCACGAGGACGACACCGAAGAGAG
CGAGAACAGATTCCGAGACGGTTGCTGCGCCGGAACCGGCAGCTCAGGCGACAGGTGAGC
GTAGGGGACGTGGGAGACCTCCGAAGGTGAAGCCGACGGTGGTTGCTCCGGTTGGGTGCT
GAATTAATCGGTACTTATGCAATTCGGAATCTTTAGTTACTGAAAAATGGAATCTCTTA
GAGAGTAAGAGAGTGCTTTAATTTAGCTTAATTAGATTTATTTGGATTCTTTTCAGTATT
TGGATTGTAACTTTAGAAATTTGTGTGTGTGTGTGTGCTTAGTCCTGAGATAAGATATAA
CATTAGCGACTGTGTATTATTATTACTGCATTGTGTATGTGAACTTTGTTCTCTT
GTTGAAAAA

>G280 Amino Acid Sequence (domain in AA coordinates: 97-104,130-137-155-162,185-192)

MAFDLHHGSASDTHSSELPSPFSLPPYPQMIMEAIESLNDKNGCNKTTIAKHIESTQQTLF
PSHMTLLSYHLNQMKKTGQLIMVKNYMKPDAPPKRGRRPPKQKQAESDAAAAAVV
AATVVSTDPPRSRRPPPKPDSEPPQEKVITGSGRPRGRPPKRPRTDSETVAAPPEPAAQ
ATGERRRGRPPKVPKPTVAVPGC*

>G3 (16..477)

GTTTGTCTTTTATCAATGGAAAGAGAACAAGAAGAGTCTACGATGAGAAAGAGAAGGCAG
CCACCTCAAGAAGAAGTGCCTAACCACGTGGCTACAAGGAAGCCGTACAGAGGGATACGG
AGGAGGAAGTGGGCAAGTGGGTGGCTGAGATTCTGTGAGCCTAACAAACGCTCACGGCTT
TGGCTTGGCTCTTACACAACCGATATCGCCGCCGCTAGAGCCTACGACGTGGCCGCTCTC
TACCTCCGTGGCCCTCCGCACGTCTCAACTTCCCTGATCTTCTCTTGCAAGAAGAGGAC
CATCTCTCAGCCGCCACCACCGCTGACATGCCCGCAGCTCTTATAAGGGAAAAAGCGGCG
GAGGTGCGCGCCAGAGTCGACGCTCTTCTAGCTTCTGCGGCTCCTTCGATGGCTCACTCC
ACTCCGCCGGAATAAAACCCGACTTGAATCAAATACCCGAATCCGGAGATATATAGTCA
ATTTATATACATGTAGTTTGTGTTTGTGATTAGAAGATTACATTTACATACAAGATACA
CATAGATACTGGAAATATAGGTATGTATACATTCAAATATCTTATGTATCAAAGAA
TTTTATAGATTCTGATTAGCTTTTGTGTTTGTGTTTGTGATAAGAACTCTGATTAGTTGTC
CGGAGACAAAACCGGCTAAGAGCAATCCATGAGAAGCTAGCGAGTGTTTTTAGTTCAAG
TTGTAATATAAATGCATATTAATTCTTTAGTAATTTTGT

>G3 Amino Acid Sequence (domain in AA coordinates: 28-95)

MEREQEESTMRKRRQPPQEEVPHVATRPYRGIRRRKWGWAEIREPNKRSRLWLGSY
TTDIAAARAYDVAVFYLRGPSARLNFDDLLQEEHLSAATTADMPAALIREKAAEVGAR
VDALLASAAPSMASHTPPVIKPDNLQIPESGDI*

>G343 (1..795)

ATGGAGCTCTATGGETTATCTTCACAGACTTACTTCGAATCGACGACCTTCTTGATTTC
TCCAACGAAGACATCTTCTCCGCTTCTTCTCCGGTGGTTCCACCGCCGCTACTTCTCT
TCTTCTTTCCCTCCTCCTCAAACCCCTAGTTTCCACCAACCACCATCTCCCTTCCCTCCGCC
GATCATCACTCCTTCTCCACGACATTTGCGTTCCAGTGATGACGACGCTCATCTTGAA
TGGCTTTTCGAATTCTGTGACGATTCTTTCGCTGATTTCGGGCGAATCCATTAGGAGGA
ACTATGACTTCTGTCAAACTGAAACTTCTTTCCGGGGAACCAAGAAGCAAACGATCA
AGAGCTCCTGCTCCTTTCGCCGGAACATGGTCTCCGATGCCACTGGAATCCGAGCATCAG
CAGCTTCACTCCGCCGCAAAATTCAAGCCAAAGAAAGAAACAATCCGGCGGAGGAGGAGGA
GGAGGAGGAAGACATCAGTCATCGTCATCGGAGACTACGGAAGGAGGAGGAATGAGGAGA
TGTACTCACTGTGCATCGGAGAAAACGCCACAGTGGAGGACAGGACCACTTGGACCTAAA

ACACTATGTAACGCTTGTGGAGTCCGGTTTAAATCCGGTAGACTTGTACCGGAATATAGA
CCGGCTTCGAGTCTACTTTTGTGTTTACTCAGCATTCAAACCTCTACCGGAAAGTGATG
GAGCTTCGACGGCAGAAAGAAGTTATGAGACAACCACAACAAGTTCAACTTCATCACCAC
CACCACCCGTTTTAG

>G343 Amino Acid Sequence (domain in AA coordinates: 178-214)
MDVYGLSSPDLLRIDLLDFSNEDIFSASSSGGTAATSSSFPPQNPFSFHHHLPSSA
DHHSFLHDICVPSDDAAHLEWLSQFVDDSFADFPANPLGGTMTSVKTETSFP GKPRSKRS
RAPAPFAGTWSMPLESEHQQLHSAAKFKPKKEQSGGGGGGGGRHQSSSSETTEGGGMRR
CTHCASEKTPQWRTGPLGPKTLCNACGVRFKSGRLVPEYRPASSPTFVLTHQHSNSHRKVM
ELRRQKEVMRQPQQVQLHHHHHPF*

>G363 (1..780)
ATGAGACCAATATTAGACCTCGAAATTGAAGCTTCATCGGGCAGTAGTAGCAGCCAAGTG
GCCTCAAACCTGTCTCCGGTTGGGGAAGATTACAAACCAATCTCGCTGAATCTTAGCCTC
AGTTTCAACAACAACAACAATAATCTGGATCTTGAATCATCGTCTTTGACGCTGCCA
CTTTCGAGCAGAGTGAGAGTAGTAACCCGAGCAGCAGCAGCAACAACAACCATCTGTA
TCAAAGAGAGTCTTCTCTTGAACCTACTGCCAAAGGAAGTTCTATAGCTCTCAAGCGCTA
GGTGGTCACCAAAACGCTCACAACGTGAGAGAACACTCGCCAAACGCGCTATGCTATGG
GTCTTGCTGGGGTCTTCCCGGTAGAGGATCAAGTAGCAATTATGCGGCTGCTGCCACAG
CAGCCGCTCTCTGTTGTTGCGCTTACCGGAAGCGGAAACGGGAACATGACATCGTTCAGG
ACTTTGGGAATCCGGGCACATTCTCGGCGCAGCAGCTCAGCATGACAAGGCAGACACCA
GAAACACTTATTAGAAACATTGCCAGGTTCAACCAGGGGTATTTGCGTAATTGTATACCT
TTTACGTGGAGACGACGAGGCGGAGATGCTCTGCGCGGGGAGTTTCCGCAAGCTACG
AATGCGGTTGCGGTTGAAGCGGTAATGATAATTTAGGTGAAAGAAAAATGGATTTCTTG
GACGTCAAGCAAGCGATGGATATGGAAGTTCTCTTCCAGATCTAACCTGAAGCTTTGA

>G363 Amino Acid Sequence (domain in AA coordinates: 87-108)
MRPILDLEIEASSGSSSSQVASNLSPVGEDYKPIISLNLSLFNNNNNNLDLESSSLTLP
LSSTSESSNPEQQQQQPSVSKRVFSCNYCQKQFYSSQALGQHNAHKRERTLAKRAMLW
VLLGSSPVEDQVAIMRLLPQQPLSCLPLHSGNGNMTSFRTLGIRAHSSAHDVSMTRQTP
ETLIRNIARFNQGYFGNCIPFYVEDDEAEMLWPGSFRQATNAVAEAGNDNLGERKMDFL
DVKQAMDMESSLPDLTLKL*

>G370 (1..774)
ATGGACGAAACCAACGACGAAGAGAAACTCAGGATTTTCATGAACGTCAACGTTGAATCC
TTCTCTCAGCTTCTTTTCATCCGCCGTACTCTCCCAAAGAAAAAGCCGCATTTATTCGT
CTCTTCGGCCAAGAGCTCGTCGGTGATAACTCCGACAACCTTATCCGCAGAACCTTCTGAT
CATCAAACCACTACCAAGAAGATGAGAGCTCTGAGAATATCAAGGACAAAGACAAAGAA
AAAGATAAGGACAAAGACAAAGATAACAACAACAACAGGAGATTTCGAGTGTCACTACTGC
TTCAGAAACTTCCCAACTTCTCAAGCCCTAGGTGGACATCAAACGCTCACAACGCTGAA
CGTCAACACGCCAAACGCGGTTCCATGACATCATACCTTCATCATCATCAGCCTCATGAC
CCTCACCATCATCTACGGCTTCTTCAACAACCACCACCACCGTCACTATCCGTCTTGGACG
ACGGAAGCTAGATCATACTACGGCGGAGGGGGACATCAAACGCCGCTCGTACTACTCAAGG
AATACTCTTGCTCTCTTCTTCTAACCACCGACAATCAACGGAAGTCTTTAGGTTTG
TGGCGTGATACCGCTTCCACGTCAACAAATACTATTAAGGCGTTTACTCATCTTACCA
GCTTCAGCGTTTAGGTGCGATGAGCAAGAGACTAATAAGGAGCCTAATAACTGGCCGTAC
AGATTGATGAAACCAATGTGCAAGATCATGTGAGTCTCGATCTTCATCTCTGA

>G370 Amino Acid Sequence (domain in aa coordinates: 97-117)
MDETNGRRETHDFMNVNVSFSQLPFIRRTPPKEKAAIIRLFGQELVGDNSDNLSEPSD
HQTTTTKNDESSSENIKDKDKEKDKDKDNNNNRRFECHYCFRNPPTSQALGQHNAHKRE
RQHAKRGSMTSYLHHHQPHDPHHIYGLNNHHHRHYPSTTEARSYGGGGHQTPSYYSR
NTLAPPSSNPPTINGSPLGLWRVPPSTSTNTIQGVYSSSPASAFRSHEQETNKEPNWPY
RLMKPNVQDHVSLDLHL*

>G385 (37..2202)
TAGGGTTTGCTTTTCAGTTTCCGGAGTATAAGAAAAGATGTTTCGAGCCAAATATGCTGCTT
GCGGCTATGAACAACGACAGCAATAACCACTACAACACGAGACAAACAATAAT
GAAGGATTTCTTCGGGACGATGAATTTCGACAGTCCGAATACTAAATCGGGAAGTGAGAAT
CAAGAAGGAGGATCAGGAAACGACCAAGATCTCTTCATCTAACAAGAAGAAACGATAT
CATCGACACACCAACTTCAGATCCAGGAGATGGAAGCGTTCTTCAAAGAGTGTCTTCAC
CCAGATGACAAGCAAAGGAAACAGCTAAGCCGTGAATTGAATTTGGAACCTCTTCAGGTC

AAATTCTGGTTCCAAAACAAACGTACCCAAATGAAGAATCATCACGAGCGGCATGAGAAC
TCACATCTTCGGGCGGAGAACGAAAAGCTTCGAAACGACAACCTAAGATATCGAGAGGCT
CTTGCAAATGCTTCGTGTCTAATTGTGGTGGTCCAAACAGCTATCGGAGAAATGTCATTC
GACGAACACCAACTCCGTCTCGAAAATGCTCGATTAAAGGAAGAGATCGACCGTATATCC
GCAATCGCAGCTAAATACGTAGGCAAGCCAGTCTCAAACCTATCCACTTATGTCTCTCTCT
CCTCTTCCTCCACGTCCACTAGAACTCGCCATGGGAAATATTGGAGGAGAAGCTTATGGA
AACAAATCCAAACGATCTCTTAAGTCCATCACTGCACCAACAGAATCTGACAAACCTGTCT
ATCATCGACTTATCCGTGGCTGCAATGGAAGAGCTCATGAGGATGGTTCAAGTAGACGAG
CCTCTGTGGAAGAGTTTGGCTTTAGACGAAGAAGAATATGCAAGGACCTTTCCTAGAGGG
ATCGGACCTAGACCGGCTGGATATAGATCAGAAGCTTCGCGAGAAAGCGCGGTTGTGATC
ATGAATCATGTTAATCATCGTTGAGATTCTCATGGATGTGAATCAATGGTCGACGATTTTC
GCGGGGATGGTTTCTAGAGCAATGACATTAGCGGTTTTATCGACAGGAGTTGCAGGAAAC
TATAATGAGAGCTCTTCAAGTATGAGCGCAGAGTTTCAAGTTCATCTCCATTAGTCCCA
ACACGTGAAACCTATTTTCGCACGTTACTGTAAACAACAAGGAGATGGTTCGTGGGCGGTT
GTCGATATTTTCGTTGGATAGTCTCCAACCAATCCCCCGGCTAGATGCAGGCGGCGAGCT
TCAGGATGTTTGATTCAAGAATTGCCAAATGGATATTCTAAGGTGACTTGGGTGGAGCAT
GTGGAAGTTGATGACAGAGGAGTTCATAACTTATACAAACACATGGTTAGTACTGGTCAT
GCCTTCGGTGCTAAACGCTGGGTAGCCATTCTTGACCGCCAAATGCGAGCGGTTAGCTAGT
GTCATGGCTACAAACATTTCTCTGGAGAAGTTGGCGTGATAACCAACCAAGAAGGGAGG
AGGAGTATGCTGAAATTTGGCAGAGCGGATGGTTATAAGCTTTTGTGCAGGAGTGAGTGCT
TCAACCGCTCACACGTGGACTACATTGTCCGGTACAGGAGCTGAAGATGTTAGAGTGATG
ACTAGGAAGAGTGTGGATGATCCAGGAAGGTCTCCTGGTATTGTTCTTAGTGACCCACT
TCTTTTGGATCCCTGTTCTCTCCAAAGCGAGTCTTTGACTTCCTCAGAGACGAGAATTCA
AGAAATGAGTGGGATATTCTGTCTAATGGAGGAGTTGTGCAAGAAATGGCACATATTGCT
AACGGGAGGGATACCGGAACTGTGTTTCTCTTCTTCGGGTAAATAGTGCAAACTCTAGC
CAGAGCAATATGCTGATCCTACAAGAGAGCTGCATTGATCCTACAGCTTCTTTTGTGATC
TATGCTCCAGTCGATATTGTAGCTATGAACATAGTGCTTAATGGAGGTGATCCAGACTAT
GTGGCTCTGCTTCCATCAGGTTTGTCTATTCTTCTGATGGTAATGCCAATAGTGGAGCC
CCTGGAGGAGATGGAGGGTCTGCTCTTGTACTGTTGCTTTTCAGATTCTGGTTGACTCAGTT
CCTACGGCTAAGCTGTCTCTTGGCTCTGTTGCAACTGTCAATAATCTAATAGCTTGCACT
GTTGAGAGAATCAAAGCTTCAATGTCTTGTGAGACTGCTTGAAAACCATCCATTAGC
>G385 Amino Acid Sequence (domain in AA coordinates: 60-123)
MFEPMMLLAAMNNADSNHNNHEDNNNEGFLRDDEFDSPNTKSGSENQEGSGNDQDPL
HPNKKRYHRHTQLQIQEMEAFFKECPHPDDKQRKQLSRELNLEPLQVKFWFQNKRTQMK
NHHERHENSRLRAENELRNDNLRYREALANASCPNCGGPTAIGEMSFDHQLRLNARL
RBEIDRISAIKAYVGPVSNYPLMSPPPLPPRPLELAMGNIGGEAYGNPNPDLKLSITA
PTESDKPVIIDLSVAAMEELMRMVQVDEPLWKSLLALDEEYARTFPRGIGPRPAGYRSEA
SRESAVVIMNHVNIVEILMDVNQWSTIFAGMVSAMTLAVLSTGVAGNYNGALQVMSAEF
QVPSPLVPTRETYFARYCKQQGDGWSAVVDISLDSLQPNPPARCRRRASGCLIQELPNGY
SKVTWVEHVEVDDRGVHNLKHMVSTGHAFGAKRWVAILDRQCERLASVMATNISSGEVG
VITNQEGRRSMLKLAERMVISFCAGVSASTAHTWTTLSGTGAEDVRVMTRKSVDDPGRSP
GIVLSAATSFWIPVPPKRVDFLRDENSRLNEWDILSNGGVVQEMAHIANGRDGTGNCVSL
RVNSANSQSNNMLILQESCIDPTASFVIYAPVDIVAMNIVLNGGDPDYVALLPSGFAILP
DGNANSAGPGDGGSLLTVAFAQILVDSVPTAKLSLGSVATVNNLIACCTVERIKASMSCET
A*
>G439 (128..967)
TATAAATCTTCGTTCTACTTTTTTTTCTTCCATAATATAGTCAATTCGTTTTCTTAATT
AGGGCTTCTTCTCTPTGTTTCTCCAATCTTTATTAGTTTATTATTATTATTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATTCATGAAGGTAACCTCATCTTCTTCTACTTCAATTCATCAAATCCAAAACC
ATTAACCTCCTAATTCATCCCTAATAATGACCAAGTCTTACCGGTATCTAACCAACCGG
TCCGATTGGGCTAAACAGCTCACTCCAACACAAATCTTCAAATTCAGACAGAGTTACA
TCTCCGGGCAAAACCAATCTCGTCTCGCTGGTAGTCATCTTCTCACCCTAAACCAAC
CTCAATGAAGAAAATCGACGTAGCAACTAAACCGGTTAAACTATACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTCCGGCTACCTAAACCAACCGGTTATG
GCTCGGTACGTTTCAAGACGGCTCAAGAAGCTGCATTAGCTTACGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTGTTTCGTCAGGACACTATAA

ACAGATATTGTCTCCGTCTATCAACGCAAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCTCAGATCGAGAAACAGAACAAAACAGAGGAGGTGCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCGGAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGCACAAGTATCCCTTTCTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTTATCTTTTGTATTTCATTTGTCTCTAAATTGTAGAATTTTATTTTC
AGAGCTTTTGAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGAACCTAACTCT
GTTTTCTTTTGAAGTATTGTTTATAATGGGCCGTTGAATGGGCCTTATTGATTTAAACA
GCCCAAGTTTTTAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)
MAMALNMNAYVDEFMEALEPFMKVTSSTSSSNPKPLTPNFI PNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPKVLYRGVRQRQ
WGKWWAEIRLPKNRTRLWLTGFETAQEAALAYDQAAHKIRGDNARLNFPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEEVLSGFSKPEKEPEFGEIYGCYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G440 (237..1301)

AAAAATCAGTTTTCATAACACGTTTTTCTCTCTCACCACCAAAAAAATCTTTTGT
TCTTGTACCAAAAAATCTCGTGATAAATCTCTCAAACTTTGTTTTATTTCTTCTTGA
TTCTCTCGAAATCTCTCTCAACAAACCCAGAACTTTCTTGATTTCGCAAGCTTTTCTTC
CTTTTATATTCTCTCATTTTGATGCGAATATAGAGAGAGTCCATAAAAGAAACAGTAATGG
ACGAATATATTGATTTCCGACCATTGAAGTACACAGAGCACAAGACTTCAATGACTAAAT
ACACCAAAAAGTCATCGGAAAACTTTCCGGTGGTAAGTCATTGAAAAAGGTTAGTATTT
GTTTACTGATCTGACGCAACAGATTCAATGACGAGACGAAGAAGATTTCTTGT
TTCTCGCCGGAGAGTCAAAAGATTCTGTTAACGAGATCACTGTTGAGCCTAGCTGTAACA
ACGTCGTCACCGGAGTTTCGATGAAAGATAGAAAGAGACTCTCTCTCTCCGATGAAA
CTCAATCTCCGGCGTCGAGTCGTCAACGTCTTAATAACAAAGTTTCAGTCTCCGGTCAGA
TAAAGAAGTTCCGTCGGTGTAGACAACCGCCATGGGGGAAATGGGCGGCGGAGATTAGAG
ATCCGGAGCAACGTCCGAGGATTTGGCTCGGGACTTTTGAGACGGCGGAGGAAGCTGCCG
TGGTTTATGATAACGCCGCTATAAGACTCCGTGGACCGGACGCTTTAACTAATTTCTCCA
TACCGCCTCAAGAAGAGGAAGAAGAAGAAGAACCGGAACCGGTTATTGAGGAGAAACCGG
TTATTATGACGACGCCAACCAACAACATCGAGTTCTGAATCAACTGAAGAAGATTTAC
AACATCTCTCATCTCTACTTCCGTTCTCAATCACCGGTCAGAAGAGATTCAACAAGTAC
AACACCGTTTAAATCAGCTAAACCCGAACCGGGTTTCAATGCACCATGTTGGCATA
CCGGGTTTAAATACCGGTTTGGTGAATCAGACGATTCTTCTTGGTACTCCGTTTC
TTGACAATTTTCAATGAATCACCACCAGAGATGTCAATATTGACCAACCAATGGATC
AAATTTTCTGTGAAAATGATGATATCTCAATGATATGTTGTTCTTGGGTGGTGAACTA
TGAACATTGAAGATGAGTTAACAAGTTCTAGTATCAAAGATATGGGTTCAACGTTTAGTG
ATTTTGTGATGATTCATTGATATCAGATCTATTAGTTGCTTAATATGATGATGAGAGTGAAG
AAGAAACCATCAAGCAATATCTATGGTGTGACTGAAAAATTTGGTGTTACTTTTTTTT
CTTTTCATAAGTTTCATGAGCTTTTTTGTCTTTTTTTTAAATAATTTATTTAGTTTGTCA
GGAGCTTGTAAACAGTTTGGGAGAAATAGTGGAAAAATAGTTAATTA

>G440 Amino Acid Sequence (domain in AA coordinates: 122-189)
MDEYIDFRPLKYTEHKTSMTKYTKSSEKLSGGKSLKKVSICTDPDATDSSSDEDEEDF
LFPRRRVKRFVNEITVEPSCNNVTVGSMKDRKRLSSSSDETQSPASSRQRPNNKVSVSG
QIKKFRGVRQRPWGKWAABEIRDPEQRRRIWLTGFETAEEAAVVYDAAIRLRGPDALTNF
SIPPQEEEEEEPEPVIIEKPVIMTPTPTSSSESTEEDLQHLSSPTSVLNRSEIIQQ
VQQPFKSAKPEPGVSNAPWHTGFNTGLGESDSSFPLDTPFLDNFYFNESPPEMSIFDQPM
DQIFCENDDI FNDMLFLGGETMNI EDELTS SSIKDMGSTFSDFDDSLISDLLVA*

>G5 (417..1421)

TTTTTTTTTTCGAATCTCCCCCTAATCTGTTGTTTCTCGCTTCTTCTCTGTTAATCATC
TGCTTTTCAAAAAGAAAGAAAAAGAAAAATTCGATTTCTGGGTTTGTGTTTGTGCATACA
GAAAAAATCAAGCTTATGAATTTGTGTTTAAATTTTTTGTGTTTAAATTTGAAAGGCAGGTT
TTTTTCAGAACGAGATCGTTTTTCAAATTTCTCTGATTTTACCTCTTTTTTCTTCTTA
GATTTTAGTGAATCGAGGGTGAAATTTTGGATTCCCTCTTTTCGGATCTACACAGAGGTT
GCTTATTTTCAAACCTTTTAGATCCATTTTTTTTTTAATTTCTCGGAAAAATCCCTGTTTC
TTTACTTTTTTATAAGTCTCAGGTTCAATTTTTTTCGATTCAAATTTTATTTTAAATG

CAGCTGCTATGAATTTGTACACTTGTAGCAGATCGTTTCAAGACTCTGGTGGTGAACCTCA
TGGACGCGCTTGTACCTTTTATCAAAAGCGTTTCCGATTCTCCTTCTTCTTCTGCGAG
CGTCTGCGTCTGCGTTTCTTACCCCTCTGCGTTTCTCTCCCTCCTCTCCCCGGTTATT
ACCCGGATTCAACGTTCTTGACCCAACCGTTTTCATACGGGTCCGATCTTCAACAAACCG
GGTCATTAATCGGACTCAACAACCTCTCTTCTTCTCAGATCCACCAGATCCAGTCTCAGA
TCCATCATCTCTTCCCTCCGACGCATCACAACAACAACCTCTTCTCGAATCTTCTCA
GCCCCAAAGCCGTTACTGATGAAGCAATCTGGAGTCCGCTGGATCTTGTTCGCTTACGGTT
CAGGTGTTCTTTCGAAGCCGACGAAGCTTTACAGAGGTGTGAGGCAACGTCACCTGGGGAA
AATGGGTGGCTGAGATCCGTTTGGCGAGAAATCGGACTCGTCTCTGGCTTGGGACTTTTG
ACACGGCGGAGGAAGCTGCGTTGGCCCTATGATAAGGCGGCGTACAAGCTGCGCGGCGATT
TCGCCCCGGCTTAACCTTCCCTAACCTACGTACATAACGATTTCACATCGGAGGCGATTTCG
GTGAATATAAACCTCTTCACTCCTCAGTCGACGCTAAGCTTGAAGCTATTGTATAAAGCA
TGGCGGAGACTCAGAAACAGGACAAATCGACGAAATCATCGAAGAAACGTGAGAAGAAGG
TTTCGTCGCCAGATCTATCGGAGAAAGTGAAGGCGGAGGAGAATTCGGTTTCGATCCGTTG
GATCTCCACCGGTGACGGAGTTTGAAGAGTCCACCGCTGGATCTTCCGCTTGTTCGGACT
TGACGTTCCGTCGACCCGAGGAGCCGCCGAGTGAACGAGACGTTCTCGTTGGAGAAGT
ATCCGTCGTACGAGATCGATTGGGATTTCGATTCTAGCTTAGGGGCAAAATAGGAAATTC
GCCGCTTGCAATGGAGTTTGTGAAATTCGATGACTGGCCCAAGAGTAATTAATTAAT
ATGGATTAGTGTAAATTTTCGTATGTTAATATTTGTATTATGGTTTGTATTAGTCTCTCT
GTGTCCGTCAGCTTGGCGTTTTCGTCAGGCTCGACCATGCCACAGTTTTCATTTTATG
TAATCTTTTCTTTTGTCTTATGTAATTTGTAGCTTCAGTTTCTTCATCTATAATGCA
ATTTTATTATGATTATGTG
>G5 Amino Acid Sequence (domain in AA coordinates: 149-216)
MAAAMNLYTCSRFSQDSGGLMDALVPFIKSVSDSPSSSSAASASAFHPSAFSLPPLPG
YYPDSTFLTQPFYSYGLDQQTGSLIGLNNLSSSQIHQIQSQIHPLPPTHNNNNSFSNL
LSPKPLLMKQSGVAGSCFAYGSGVPSKPTKLYRGVRQRHWGKWWAEIRLPRNTRLWLGT
FDTAEEAALAYDKAAYKLRGDFARLNFNPNLRHNGFHIGGDFGEYKPLHSSVDKLEAICK
SMAETQKQDKSTKSSKKREKVVSPDLSEKVKABENSVSIGGSPVTEFEESTAGSSPLS
DLTFADPEPPQWNETFSLEKYPSEIDWDSILA*
>G550 (1..1374)
ATGGCTGATCCGGCGATTAAGCTCTTTGGAAGACGATTCCCTTACCTGAGCTTGGTGT
GTTGATTCTTCTTCTAGCTATACCGGATTTTAAACCGAAACTCAGATTCTGTTCCGGTTA
TCAGATTCTGTGTACCGCGATGATGATGATGAAGAGATGGGTGATTCCGGTTTAGGACGA
GAAGAAGGTGATGATGTTGGTGATGGTGGAGGAGAGCGAGACTGATAAAAAGGAAGAA
AAAGATAGTGAGTGTGAGGAAGAGTCATTGAGGAATGAATCTAATGATGTTACTACTACT
ACATCGGGTATAACTGAAAAACCGAAACAACAAAAGCTGCAAGACGAATGAAGAGTCA
GGTGGTACTGCTTCTCTCAAGAGGGGAAGTTAAAGAAACCTGATAAGATTCTACCGTGT
CCGCGATGTAACAGCATGGAACCAAGTTCTGTTACTACAACAACTATAATGTTAACCAA
CCTCGCCATTTCTGCAAGAAATGTGAGAGATATTGGACAGCTGGTGGAAACGATGAGGAAT
GTTCCGGTTGGTGCTGGGAGACGTAAGAATAAGAGTCCAGCTTCTCATTATAACCGTCAT
GTAAGTATAACATCTGCGGAAGCTATGCGAAGGTGGCGAGAACTGATCTTCAACATCCT
AATGGTGCAAACTCTTCTCACTTTTGGCTCTGATTCTGTGCTTTGTGAATCTATGGCTTCT
GGATTGAATCTTGTGAGAAGTCATTGTTGAAGACACAACTGTATTGCAAGAACCCAAT
GAAGGCTTGAAGATTACCGTTCCGTTAAACCAGACAAACGAAGAAGCTGGAACAGTCAGC
CCGTTACCAAAAGTTCCATGCTTTCCAGGACCACCACCAACTTGGCCTTACGCTTGAAC
GGAGTTTTCGTGGACGATTTTACCGTTTACCCTCCACCGGCTTACTGGAGCTGCCCGGGG
GTTTCACCGGGGGCATGGAACAGCTTCACATGGATGCCACAACCAATTACCATCTGGT
TCCAATCCAAATTCCTTACACTAGGTAAACATTACGTGACGAGAACGCTGCTGAACCA
GGAACCGCTTTTGTGAAACCGAGTCACCTGGTAGGGAGAAAAGCAACCCGAGAGATGC
TTGTGGGTTCCCAAGACGCTGAGGATTGATGATCCAGAGGAAGCTGCTAAAAGTTCCATC
TGGGAAACATTAGGGATCAAAAAGACGAAAATGCGGATACTTTCCGAGCTTTTCAGATCA
TCAACCAAGAAAAAGCAGTCTTTCTGAAGGAAGACTTCCGGGAAGAAGACCGGAGTTG
CAAGCGAATCCTGCTGCTTTCTAGGTGAGCAAACTTCCATGAGAGCTCATAG
>G550 Amino Acid Sequence (domain in AA coordinates: 134-180)
MADPAIKLFGKTIPLPELGVDSSSSYTGFLETETQIPVRLSDSGDDDEEMGDSGLGR
EEGDDVGGGGESETDKKEKDESECQESLRNESNDVTTTTSGITEKTETTTKAAKTNEES
GGTACSQEGKLLKPKDILPCPRCNSMETKFCYNNYNVNQPRHFCCKKCQRYWTAGGTMRN

VPVAGARRKNSPASHYNRHVSITSAEQMKVARTDLQHPNGANLLTFGSDSVLCESMAS
GLNLVEKSLKLTQTVLQEPNEGLKITVPLNQTNEEAGTVSPLPKVPCFPGPPPTWPHYAWN
GVSWTILPFYPPPAYWSCPGVSPGAWNSFTWMPQPNPSGSPNPNSTLGKHSRDENAAEP
GTAFDETESLGREKSKPERCLWVPKTLRIDDP EAAKSSIWETLGIKKDENADTFGAFRS
STKEKSSLSEGRLPGRPELQANPAALSRSANFHES*

>G670 (28..1152)

CACAGCATTGCAGCTGTGAATAACTAAATGGGGAGACATTCTTGCTGTTACAAACAAAAG
CTGAGGAAAGGGCTTTGGTCTCCTGAAGAAGACGAGAAGCTTCTTACTCACATCACCAAT
CACGGCCATGGCTGCTGGAGCTCTGTCCCTAAACTCGCTGGTTTGACAGAGATGTGGGAAG
AGTTGTCGACTCGAGCAGATCTGGTACCGCCGACTAAGATGGATCAATTACTTGAGACCT
GATTAAAGAGAGAGCTTTTTCTCTGAAGAAGAGAATCTCATCGTCAAGCTTCTATGCC
GTCCCTGGGAAACAGATGGTTCACAGATTGCGTCAAGCTTCCGGGTAGAACCGACACGAG
ATCAAGAATCTATGGAATCAAGCATCAAGAAGAACTGAAACAAAGAGGCATTGACCCA
AACACACACAAGCCCATCTCTGAAGTGGAGAGTTTTAGCGACAAAGACAAACCAACAACA
AGCAACAACAAAAGAGCGGTAACGATCACAAGTCTCCTAGTTCCTTCTGCGACTAAC
CAAGACTTCTTCTCGAAAGGCCATCTGATTTATCCGACTACTTCCGATTTTCAAGGCTT
AACTTCAACTCCAATCTAGGACTCTCTGTTACAACGATTCTTCACTCTGCTCGATGATT
CCGCCGCGAGTTAGTCCCGGGAACATGGTTGGTTCTGTCTTCAGACACCGATATGCGTA
AAGCCTCGGATAGTCTTCTCCCGACAACAACAGTTCGAGTCTTATCTCCGGAGGAGAT
CATGTGAAATTTGGCTGCACCAAACTGGGAATTTGAGACAAACAATAATACCTCAAAT
TTCTTCGACAATGGCGGATTCTCATGGTCTATCCCAAATTTCTTACTTCTTCTTCAAA
GTCAAACCAATCATAACTTCAAGAAATAAAATGGTCAGAGTATTTGAACACACCGTTC
TTCATAGGGAGTACTGTACAGAGTCAAACCTCTCAACCAATCTACATCAAATCAGAAACA
GATTACTTAGCCAATGTTTCAAACATGACAGATCCTTGGAGCCAAAACGAGAACTTGGGC
ACAAGTGAAGTACTGACGTGTTCTCAAGGATCTTCAAGAGATGGCGCTCTCTTTTGGT
CAGTCCCTTTAGCTTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCTTTCT
AGATATACAAATACATACAATGTCAATACGTACAGTGGATTTAAGTGTCTGTATATTTCT
ATGGGCGAGCTGTCTTTATTTTATGTTTAAAAA

>G670 Amino Acid Sequence (domain in AA coordinates: 14-122)

MGRHSCCYKQKLRLKGLWSPEEDEKLLTHITNHGHGCVSSVPLAGLQRCGKSCRLEQIWI
RRLRWINYLRPDLKRGAFSPPEENLIVELHAVLGNRWSQIASRLPGRDNEIKNLWNSSI
KKLKRQIDPNTHKPISEVESFSDKDKPTTSNNKRSNDHKSPSSSSATNQDFFLERPS
DLSDFGFGFKLNFNSNLGLSVTTDSSLCSMIPPQFSPGNMVGSLQTPVCVKPSISLPPD
NNSSSPISGGDHVKLAAPNWEFQTNNTSNFFDNGGFSWSIPNSSTSSSQVKPNHNFEE
IKWSEYLNTPFFIGSTVQSQTSQPIYIKSETDYLANVSNMTDPWSQENLGTETSDVFS
KDLQRMVVSFGQSL*

>G760 (175..1878)

TGCTTAATTCGAATGCCATCGTGATCGATTCTCTCTCTCTCTCTCTTCCAATTTCCCA
ATTCTTTTTTAAACCCTAATTTTTCAGATATCTGATTATCTCTTGATTTCTTCTACTC
GATTTGCTCCCATATAAAACCCCTTACTTTCTTCAAGTTCTGGTTTTCCACCGATTGATGGT
CGTGGCTCAGTGACGTCGCTTGCTCCTGGGTTCCGTTTTTCAACCGACGGATGAGGAAGT
GTTGCTACTACCTTAAGCGTAAGGTCTGCAACAAACCCCTTAAAGTTCGATGCTATTTCC
GTCACCGACATATACAAGTCTGAGCCTTGGGATCTACCAGATAAGTCAAGCTGAAAAGT
AGAGACTTGGAAATGGTACTTCTTTAGTATGCTGGATAAGAAGTACAGTAATGGTTCCAAG
ACGAATCGTGCTACGGAGAAAGGGTATTGGAAGACGACTGGGAAAGATCGGGAGATTTCGT
AATGGTTCAAGAGTCTTTGGGATGAAGAAGACACTTGTTTATCACAAGGTCGAGCTCCT
CTTGGTGAAGGACCAATTTGGGTTATGCTGAGTATCGGCTTTCTGATGAGGACTTGAAG
AAAGCTGGTGTGCCACAAGAAGCATATGTGTTATGTAGGATATTCCAGAAAAGTGGTACG
GGTCTTAAGAATGGGGAGCAGTATGGTGTCTCTTATCTTGAGGAGGAGTGGGAAGAAGAT
GGAATGACTTATGTACCTGCTCAAGATGCTTTCAGTGAAGGATTGGCTTTGAATGATGAT
GTTTATGTGATATTGATGACATTGACGAGAAGCCCGAAAATCTGGTGGTCTATGATGCC
GTTCTATTCTACCTAACTATTGTGATGGGGAATCAAGTAACAATGTTGAATCAGGCAAT
TACTCAGACTCTGGAATTTACATTCAACCAGGAAACAATGTTGTGACTCTGGTGGGTAC
TTTGAACAACCAATTGAACTTTTGAAGGAAGATCGGAAGCCTATTATACGGGAGGGTAGC
ATTGAGCCTTGTCTCTGTTTCCAGAGGAACAAATTGGCTGTGGTGTGCAAGACGAAAAT
GTGGTGAATCTGGAATCTTCCAACAATAATGTGTTTGTAGCTGATACATGCTACAGTGAC
ATTCCTATTGATCATACTATTACCCGATGAGCCATTCTGATGATCCTAATAACAATCTT

CCACTCAACGATGGTCTGTACCTGGAAACGAATGATCTCAGCTGTGCTCAACAAGATGAT
TTTAACCTCGAAGATTATCTCAGCTTCTTTGATGATGAGGGTTGACTTTTGACGATTCT
CTATTAATGGGACCTGAAGATTTTCTTCCCAACCAAGAAGCCCTTGACCAGAAACCTGCC
CCTAAAGAATTGGAGAAGGAGGTGCGCAGGAGGCAAAGAGGCAGTGGAGGAAAAGGAAAGT
GGCGAAGGATCTTCTTCAAAACAAGATACAGATTTCAGGACTTTGATTACAGCTCCGAAG
TACCCATTCTCAAAAAGACGAGCCACATGCTTGGAGCCATTCTACTCCATCTTCATTT
GCTTCACAGTTCCAAACAAAGGACGCAATGCGTCTACACGCAGCACAATCTTCTGGTTCA
GTTACGCTGACTGACAGGTATGATGAGAATATCAAACATGACTCTAGCAGCGGACAGCGGT
ATGGGCTGGTCAATGACAAGAACGGTAACCTCAACGTAGTCCTTTTATTGGGGTAGTC
CAACAGGATGATGCGATGACTGCCTCGGGAAGCAAGACAGGAATTACGGCGACAAGAGCT
ATGTTAGTCTTTCATGTGTTTATGGGTTCTCTACTCTCTGTTAGCTTCAAAATAGTAACC
ATGGTGTCTGCTCGGTAATAGGATCAAAGTTGAATCGTCTCAAAGACTTTTCTGGTGT
TGTACCTCTCAATCATATAGCCTTTAACTTTGGCAGTGCCTTGCTGCTCAATATTTAA
TTTTAAAAAAAAAAAAAAAAA
>G760 Amino Acid Sequence (domain in AA coordinates: 12-156)
MGRGSVTSAPGFRFPHPTDEELVRYLKRKVCNKPFDKDAISVTDIYKSEPDLPDKSKL
KSRDLEWYFFSMLDKKYSNGSKTNRATEKGYWKTGKDREIRNGSRVVMKKTLVYHKGR
APRGERTNWMHEYRLSDEDLKKAGVPQEAAYVLCRIFQKSGTGPKNGEQYGAPYLEEWE
EDGMTYVPAQDAFSEGLALNDDVYVDIDIDEKPENLVVYDAPVILPNYCHGESNNVES
GNYSDSGNYIQPNNNVVDSSGYFEQPIETFEEDRKPIIREGSIQPCSLFPFEEQIGCGVQD
ENVNLESSNNNFVADTCYSIDIPIDHNYLPDEFMDPNNNLPLNDGLYLETNDLSAQQ
DDFNFDYLSFFDDEGLTFDDSLMGPEDFLPNQEAALDQKPAKKELEKEVAGGKEAVEEK
ESGEGSSSKQDTFDKDFDSAPKYPFLKKTSHMLGAIPTSSFASQFQTKDAMRLHAAQSS
GSVHVTAGMMRLSNMTLAADSGMGWSYDKNGNLNVVLSFGVVQDDAMTASGSKGITAT
RAMLVFMCLWVLLSVSFKIVTMVSAR*
>G831 (92..1987)
TTCTTTTCATCGTGTGTCTATTATAAATATATGTCAATTGGTTTCTAAAAATTCTACC
ATTGATTGATTGATTTTTTTTCTTTAAGAGATGAATTTATTTACAAGAATCTCATCTCG
GACTAAGAAGGCCAATCTTTACTACGTAACCCCTAGTTGCTCTTCTCTGCATCGCTAGCTA
CCTTCTCGGTATTTGGCAAAACACGGCGGTTAATCCACGCGCCGCTTCGATGATTGAGA
CGGTACACCGTGCGAGGGATTACACAGACCTAATTCTACGAAAGATCTCGACTTCGACGC
GCATCACAACATTCAGATCCACCTCCGGTGACGGAACCGCCGTAGTTTCCCGTCGTG
TGCCGCGCGCTTGAGCGAGCACACGCCATGCGAAGACGCGAAGCGATCGTTGAAATCTC
GAGGGAGAGATTGGAGTATAGGCAAAAGGCATGTCCCGAGAGAGAAGAAATCTTGAAGTG
CAGAATTCCGCGCGCTACGGTTACAAAACGCGCTTCCGATGGCCGCGAGTCGTGACGT
GGCGTGGTTCGCTAATGTGCCTCACACGGAGCTTACGGTTGAGAAAAAGAATCAGAATTG
GGTCCGGTACGAGAATGATCGGTTTTGGTTCCCTGGTGAGGTACGATGTTTCCACGTGG
CGCTGATGCTTACATTGATGATATCGGACGGTTGATTGATCTCAGCGACGGCTCTATCCG
TACAGCCATCGATACCGGTTGCGGGGTGGCTAGCTTCGGTGATATCTTTTATCAAGAAA
CATTACAACGATGTCATTTGCACCAAGAGACACACGAAGCTCAAGTCCAGTTCGCACT
CGAGCGTGGTGTGCGGCGGATGATCGGAATCATGGCTACAATCCGCTACCGTACCTTC
TAGAGCCTTTGATTTAGCACATTGCTCTCGTTGCCCTTATTCCGTGGGGCCAAACGATGG
GGCTTACTTGATGGAGGTGGATAGGGTTTTAAGACCAGGAGGGTACTGGATACTTTCTGG
ACCGCCGATTAATTGGCAGAAACGGTGAAAGGGTGGGAACGGACCATGGATGATTGAA
TGCAGAGCAGACTCAGATCGAGCAGGTGCGGAGAAGCTTGTGTTGGAAGAAAGTTGTTCA
AAGAGATGATCTTGCTATTGGCAAAAACCCCTTTAACCACATTGACTGTAAGAAAACAG
AGAGGTTTTGAAAAATCCGGAGTTTTGTCTCATGATCAAGATCCCGACATGGCCTGGTA
TACGAAGATGGATTETTTGTTGACACCATTACCTGAAGTTGATGACGCTGAGGATCTAAA
GACGGTGGCCGGAGGGAAGGTAGAAAAGTGGCCGGCTAGATTAAACGCGATTCTCCGAG
AGTAAACAAAGGCGCTCTCGAGGAAATCACACCTGAAGCTTTCTGGAGAACACGAACT
GTGGAACAGAGAGTTTCTTATTACAAGAAGTTAGATTACAGTTGGGTGAAACGGGAG
ATACAGAACTTAGTCGACATGAACGCTTACCTCGGTGGATTTCGCGGCGGCTCTAGCGGA
TGATCCGGTCTGGGTGATGAACGTTGTCCCGTGGAGGCTAAGCTCAATACGCTCGGTGT
CATCTACGAGCGTGGTCTAATCGGAACGTATCAAACTGGTGTGAAGCCATGTCGACGTA
TCCAAGAAGTATGATTTTATCCATGCTGACTCGGTTTTTACATTGTACCAAGGTCAATG
TGAACCGGAGAGATATTGTTGGAGATGGACCGAATTCTTAGACCGGGTGGTGGTGTGAT
TATAAGAGATGACGTGGACGTTTTGATCAAGGTTAAGGAATTAACCAAGGATTAGAATG

GGAAGGTAGAATTGCTGACCACGAGAAGGGTCCTCATGAAAGAGAGAAGATTACTATGC
GGTGAAACAGTATTGGACCGTTCTGCGCCTGATGAAGATAAAAAACAACACTAGTGCTCT
CTCCTGATTTTTGAGTTTTTTTTTCTTACAATGTTTTTTTTTTTTTTTCAATTTTT
TATACAACAATAAATTCTCAATAATTGTTGTCGCGGCCG

>G831 Amino Acid Sequence (domain in AA coordinates: 470-591)
MNLFTRISSRTKKANLYVTLVALLCIAASYLLGIWQNTAVNPRAAFDDSDGTPCEGFTRP
NSTKOLDPDAHNIQDPPPVTTETAVSFPSCAAALSEHTPCEDAKRSLKFSRERLEYRQRH
CPEREELKCRIPAPYGYKTPFRWPASRDVAWFANVPHTELTVKKNQNWVRYENDRFWF
PGGGTTFPRGADAYIDDIGRLIDLSDGSIRTAIDTGCQVAFSGAYLLSRNITMSTFAPRD
THEAQVQFALERGVPMIGIMATIRLPYPSRAFDLAHCSRCLIPWQNDGAYLMEVDRVL
RPGGYWILSGPPINWQKRWKGWERTMDDLNAEQTQIEQVARSLCWKKVVQRDDLAIWQKP
FNHIDCKTREVLKNPEFCRHDQDPMAWYTKMDSCLTPLPEVDDAEDLKTAVAGGKVEKW
PARLNAIPPRVNGALEBITPEAFLENTKLWKQVSYKKLDYQLGETGRYRNLVDMNAY
LGGFAAALADDPVWVMNVVPVEAKLNTLGVIYERGLIGTYQNWCEAMSTYPTYDFIHAD
SVFTLYQGQCEPEEILLEMDRILRPGGGVIIRDVDVLIKVKELTKGLEWEGRIADHEKG
PHEREKIYAVKQYWTVPAPDEDKNNTSALS*

>G864 (503..1534)
TGCAAAAACATTTTCTGTCTCTCCTCTGCCCAAATTTTTTTCTTTCCAGGAATATTTCT
CTAGAAAAACCAAGCAAAGCTTTAACCCTTCTCTCCTCCAAAAGTAGCATCTTCTCTT
TTTCTATTTCTCCTTTCTCTTCTTATCTCTCTCTCGTTGTGAACGATTCCTTAAGAAAT
ATAACCAAAGCCCTTTTCTCCTTTCTCAACTTTCCGGGAAAAATCTTCACGCAGCAAG
GTTTCTCTCTCGGCTCTCGCAGTGTTTTTCGGGCCTTTGTCTTTCTATAAAAAAAAAA
TTTCGCGTCTTTAAGAAAATTTTCCACCTAGAGAAGAAGAAGAGTATCACTCTTGTG
TTCAAGTTTCTCTCTTTAATAAAAAATCCATCTTTATCTTTGTCTTTCTCTTTTGTG
TTTCCCTAATCTCTATGTTATAAACACACAGAGAGAAAAAGTACAGTCTCGAGTCAA
AAACAGAGAAATACGAAAGAAAAATGGAAGCGGAGAAGAAAAATGTTTCTACCGAGAATCAA
ATTACAGAGACAAAAACACACGACAACAATCGTATCGGAGTTAAACCAACTCACCA
AACCAGGATTCTTCGTATCTCAGTCACTGACCCAGACGCTACTGATTCTCCAGTGACGA
CGAAGAAGAAGAACATCAACGCTTGTCTCTAAACGCCGTCGTGTTAAGAAGTTTGTCAA
CGAAGTCTATCTCGATTCCGGTGCTGTTGTACTGGTAGTTGTGGTCAAATGGAGTCGAA
GAAGAGACAAAAGAGAGCGGTTAAATCGGAGTCTACTGTTCTCCGGTGTGTTTCAGCGAC
GACGACTACGACGGGAGAGAAGAGTTCCGAGGAGTGAGACAGCGTCCATGGGGAAAATG
GGCGGCGGAGATAAGAGATCCGTTGAAACGTGTACGGCTCTGGTTAGGTACTTACAACAC
GGCGGAAGAAGCTGCTATGTTTACGATAACGCCGCTATTACGCTTCGTGGTCCCGACGC
TCTGACTAATTTCTCAGTCACTCCGACAAACAGCGACGGAGAAGAAAGCCCAACCGTC
TCCGGTGAAGAAGAAGAAGAAAAACAACAAAAGCAAAAAATCCGTTACTGCTTCTTC
CTCCATCAGCAGAAGCAGCAGCAACGATTGTCTCTGCTCTCCGGTGTCTGTTCTCCGATC
TCCTTTTCGGCTCGAGCAATTTCTCCGGCATTTCTTCATCACCAGTCGCGGCCGTTGTAGT
CAAGGAAGAGCCATCCATGACCAACGGTATCTGAAACTTTCTCTGATTTCTCGGCGCCCTT
GTTCTCAGATGATGACGTGTTTCGATTTCGGGAGCTCAGTGGTCCCGACTATCTCGGCGG
CGATTTATTTGGGAAGATCTATTACGGCGGATATGTGTACGGATATGAATTCGGATT
CGATTTTCGGATCCGGATTATCCAGCTGGCACATGGAGGACCATTTTCAAGATATCGGGGA
TCTATTCCGGTCCGATCCTCTTTAGCTGTTAATAATATTTTAAATAAATAAATAGTTA
TACCGGCCGTTACTAAACGGAACCGGAGAAAGTTTGTATACCGGTGACATAAAATCTCG
GTTATGTTTCGTAATCTTTTTTCTTTGTTATATATAAAAAATGAATGAACTGAATTAA
TGTAAGTTAATGTTGATAATTATTACGTTTAAAGTTTGAIAAAAAAAAAAAAAAAAAA
AAAAAA

>G864 Amino Acid Sequence (domain in AA coordinates: 119-186)
MEAEKMMVLPRKIFTEHKINTTIVSELINTHQTRILRISVTDPDATDSSSDDEEEHQR
FVSKRRRVKFKFVNEVYLDGAVVTGSCGQMESKKRQKRAVSESTVSPVVSATTTTTGK
KFRGVRQRPWGKWAAEIRDPLKRVRLWLGTYNTEEAAMVYDNAAIQLRGPDALTNFSVT
PTTATEKKAPPPSPVKKKKKKNNKSKSVTASSSISRSSNDCLCSRVSVLRSPPFAVDEF
SGISSSPVAADVVEEPSMTTVSETFSDFSAPLFSDDDVDFRSSVVPDYLGGDLFGEDL
FTADMCTDMNFGPDFGSLSSWHMEDHFQDIGDLFGSDPLLAV*

>G884 (31..1575)
TTTTTTTTTTGTTTGTAAATTTGGGGATCGATGTCGGAAAAGGAAGAAGCTCCGTTCGACA
TCGAAGTCCACCGAGCTCCGTCGCGTCCGACTTTATCTCTTCTCCACGGCCGTTTAGT

GAGATGTTCTTTAACGGTGGCGTTGGATTGAGTCTGGTCCGATGACTCTGGTCTCTAAT
ATGTTCCCTGATTCCGATGAGTTTAGGTCTTTCTCTCAGCTTCTCGCTGGAGCCATGTCT
TCTCCAGCGACTGCAGCTGCTGCTGCTGCTGCTGCGACGGCTAGTGATTACCAGAGACTT
GGTGAAGGGACTAATAGCTCTAGTGGTGATGTTGACCCGAGATTCAAGCAAAACAGACCA
ACCGGTTTGATGATTTCTCAATCTCAATCGCCGTCGATGTTACCGTACCGCCTGGTTTA
AGTCCAGCTATGTTGCTCGATTACCAAGCTTTTGGGTCTTTCTCTCCCGTTCAGGGA
TCATATGGAATGACACATCAGCAAGCTCTAGCTCAAGTCACTGCTCAAGCAGTTCAAGCC
AATGCCAATATGCAACCACAAACAGAGTACCCTCCTCCCTCTCAAGTTCAATCATTTTCA
TCGGGTCAAGCGCAGATCCCGACCTCGGCTCCACTACCAGCTCAAAGAGAAACCTCAGAT
GTAACCATCATAGAGCACAGGTACACACAGCCTCTAAATGTTGACAAACCAGCTGATGAT
GGCTATAACTGGCGAAAATATGGGCAAAAGCAAGTTAAAGGTAGCGAGTTTCCACGAAGC
TATTACAAGTGATTAATCCAGGATGTCTGTCAAGAAGAAGGTTGAGAGATCTCTTGAT
GGACAAGTAACGGAGATTATCTACAAAGGTGAGCACAATCATGAACCTCCTCAAAACACT
AAGCGAGGTAACAAAGATAACACCGCAATATAAATGGGAGTTCCGATAAATAACAATCGC
GGAGTTCTGTAATTGGGGGCATCAGTTTCAAATAAGTCTCAACAAGACTAAGAGA
GAGCAACATGAAGCAGTAAGTCAAGCTACGACACACAGAGCACTGTCTGAGGCAAGTGAC
GGTGAAGAAGTTGGTAAATGGAGAACTGATGTGAGAGAGAAAGATGAGAATGAGCCTGAT
CCCAAGAGAAGAAGTACAGAAGTTCCGATTTTCAAGACAGCTCTGCTGCTTCACATAGA
ACTGTGACAGAGCCTAGAATTATTGTCCAAACGACGAGTGAAGTTGATCTTCTAGATGAT
GGATATAGGTGGCGTAAATATGGACAGAAAGTTGTCAAAGGGAATCCTTATCCGAGGAGC
TACTACAAGTGCACAACACCGGATGTGGTGTGAGGAAACATGTAGAGAGAGCAGCAACA
GATCCAAAAGCTGTAGTAACAACATATGAAGGAAAACATAACCATGACCTTCCCGCTGCT
AAATCAAGCAGCCATGCCGCTGCAGCGGCACAGTTAAGGCCAGATAATCGACCTGGCGGT
TTGGCTAACTTAAATCAACAGCAGCAGCAACAGCCGTTGCGCGCTAAGGCTTAAAGAA
GAGCAACAACCTTGAGAGAAGAAACTCTTGACCGTTTTTTCATTACAAAGCTTTCAAAT
TCCACTCACACACTTGTCTGAAAAATCTAGCAGTTTGCAGGAAAGAAACAGCTTCAAGAG
GTTGTAGTTCTTCTATGTTCTGGTGTAAACTTAAAGCTTTTAGGGTTTTTCAAGATTTCT
TGTTTACTAATACTGTATGTGAATCTTTTGTACATGAGGAAGAAAATTACAGGGGGATA
TTTTGTGTGTATCTTTGTGTTATTGTTTCAGTAAAAGATAGGTCTTACATTTTGTGTA
AAAAAAAAAAAAAAAAAAAA

>G884 Amino Acid Sequence (conserved domain in AA coordinates:227-285, 407-465)

MSEKEAPSTSKSTGAPSRPTLSLPPRFSEMFNNGVGFSPGPMTLVSNMFPDSDEFRS
FSQLLAGAMSSPATAAAAAAAAAATASDYQRLGEGTNSSSGDVDPFRKQNRPTGLMISQSQS
PSMFTVPPGLSPAMLLDSPSFLGLFSPVQGSYGMTHQALAQVTAQAVQANANMQPQTEY
PPPSQVQSFSSGQAIPTSAPLPAQRETSVDVTIEHRSQQPLNVDKPADDDGYNWRKYGQK
QVKGSEFPRSYYKCTNPGPCPVKKKVERSLDGQVTEIIYKGQHNHEPPQNTKRGNKDNTAN
INGSSINNNRGSSELGASQFQTNSSNKTREQHEAVSQATTEHLSEASDGEEVNGETD
VREKDENEPDPKRRSTEVRRISEPAPAASHRTVTEPRIIVQTTSEVDLLDDGYRWRKYGQK
VVKGNPYPRSYYKCTTPGCGVRKHVERAATDPKAVVTTYEGKHNHDLPAKSSSHAAAAA
QLRPDNRPGGLANLNQQQQQPPVARLRLKEEQTT*

>G898 (161..772)

GAAGAAAGATTCAAAAACCCTAGATTTCAAAAATCGATTGGCTGTCAAATTTCTCTCC
GGCGATTTTCTCGAGTGAAATTCGGCTCAAGGTGATTATAGCGATCATCGAATCAAAT
GATTGAAGAGGTACAAAGGTTAGTTACTTTGAGCTGAAAGATGAACACGTCAGAGGTGAG
AGTACCTCGAGGAAATCGACGGAGGAAAGCTGTGATTGATCTGAATGCGGTACCTGTTGA
TCAAGAAGGGACCTCTGCTTCTGTAGAACTCTTACGGTGCCTATTACACCGTCTCAGCC
TGCTCCTACGATGATTGATGTCGATGCTATTGAGGATGATGTTATTGAATCATCCGCTAG
TGCTTTTGCTGAAGCTAAAAGCAAATCAAGAAATGCACGTCGGAGACCTTTGATGGTTGA
TGTAAGTTCAGGAGGTACGACTAGATTCCCTGCCAACATAAGCAACAAACGAGAGGAT
TCCTTCTAGTGAATCTGTGATCGACTGTGAGCATGCCTCTGTAAATGATGAAGTCAACAT
GTCTTCGAGAGTGTCTAGATCAAAGGCTCCAGCTCCTCCACCAGAAGAGCCAAAGTTTAC
ATGTCCAATCTGATGTGTCCCTTTACGGAGGAGATGTCAACCAAGTGGGTACATCTT
CTGCAAGGGATGTATAAAGATGSCAATATCTCGCCAGGGCAAATGCCCTACTTGTAGGAA
AAAGGTTACTGCAAAAGAGCTGATTGAGTTTCTTCCAACCACTAGATGAGTGGTCCG
GCAACATCACCAGCCACCCTGTCTAATGGTTTATCAGACTATCCTCCTATTCACTTTGGA
ACATTGAAGGGACTTCGTTGACTTGGTATTTTGAATATTTTGCTTTGTGGAAGAGAAA
TATTCAGTGATCAAGAAGCCAGAAGGCCCTATCATTGATGGATATCATTTGGTAATAACT

CTTTGTTTTAGTTGTTGTTCTATGTAATTTAGGTCTCTGCAAACCTCTCAGTCGATACT
CTTCTCTCTTGATAGATGATAAGATATATGGAAAAAAAAAATTAATATTGAATCTTTACTA
AAA

>G898 Amino Acid Sequence (domain in AA coordinates: 148-185)
MNTSEVRVPRGNRRRKAVIDLNAVVDQEGTSASVRTLTVPITPSQPAPTMIDVDAIEDD
VIESSASAFEAESKSRNARRRPLMVDVESGGTTRFPANISNKRRIIPSSSEVIDCEHAS
VNDEVNMSSRVRSKAPAPPEEPKFTCPICMCPFTEEMSTKCGHIFCKGCIKMAISRQG
KCPTCRKKVTAKELIRVFLPTTR*

>G900 (1..648)

ATGGGGAAGAAGAAGTGCAGATTATGTTGTGGTGTAGCGAGAATGTATTGTGAGTCAGAT
CAAGCGAGTTTATGTTGGGATTGTGACGGTAAAGTTCACGGAGCTAATTTTCTGGTGGCG
AAACACATGCGTTGTCTTCTATGTAGCGCGTGTGAGTCACACACGCCCTGGAAAGCTTCT
GGGCTGAATCTTGGCCCAACTGTTTCTATCTGTGAGTCTTGTTTAGCTCGTAAGAAGAAT
AACACAGCTCCCTCGCCGGGAGGGATCAGAATCTTAACCAAGAAGAAGAGATCATTGGT
TGTAACGACGGAGCTGAGTCTTATGATGAGGAAAGCGATGAGGATGAAGAAGAAGAA
GTGGAGAATCAGGTTGTTCCGGCTGCGGTGGAGCAAGAAGTTCGGTGGTGAATTCGTCG
TCTTCGGTTAGTAGTGGTGAAGGAGATCAGGTGGTGAAGGACGAGACTTGATTGGAT
CTTAACCTCTCCGATGAGGAGAACCAATCTAGACCATTGAAAAGATTATCGAGAGACGAA
GGTTTGTCAAGATCAACTGTTGTGATGAATAGCTCAATCGTGAAATTACACGGAGGGAGG
AGAAAAGCAGAGGGATGTGATACATCATCGTCTCTTCTGTTTATTGA

>G900 Amino Acid Sequence (domain in AA coordinates: 6-28, 48-74)
MGKKKCELCGVARMYCESDQASLCWDCDGKVHGANFLVAKHMRCLCSACQSHTPWKAS
GLNLGPTVSIACESLARKKNSSLAGRDQNLNQEEI IGCNDGAESYDEESDEDEEEEEE
VENQVPAAVEQELPVVSSSSSVSSGEGDQVVKRRLDLNLSDEENQSRPLKRLSRDE
GLSRSTVVMNSSIVKLHGRRKAEGCDTSSSSSFY*

>G913 (108..806)

CATTCAAAAACATCATATATATACACAAACACACTTTGATACAACAAAAAACAGAAC
ACAAACAAAAACACATTGTAACATTAGTTTAAAGCATTAGCTTCTTTATGTGCAATAATA
ATAATTCTCCGACCACCGTGAATCAAGAAACGACGACGTCTCGTGAAGTCTCAATCACAT
TGCTTACTGATCAATCTCCTCAAACCTCACCAGGATCATCTTCTTCTTCAACCGAGAC
CTTCCGGTGGATCACCGGCGAGAAGAACGGCGACTGGATTATCCGGCAAGCACTCTATTT
TCAGGGGGATTGCACTACGTAACGGAATGGGTATCGGAGATTAGAGAGCCACGTA
CGACAAGAATTTGGCTCGGGACTTATCCGGTACCGGAGATGGCTGCCGCCGCTTACGACG
TGGCTGCGTTAGCTTTAAAGGACCCGACGCCGTTTGAATTTTCTGTTTAGCTTTGA
CTTACGTGGCTCCGGTTTCAAACCTCTGCTGCGGATATAAGAGCGGCTGCTAGTAGAGCAG
CGGAGATGAAGCAACCGGATCAGGGTGGGGATGAGAAGGTATTGGAACCGGTTCAACCCG
GCAAAGAGGAAGAATTAGAAGAAAGTGTCTGTAACCTCGTGTTCGTTGGAGTTTATGGATG
AGGAAGCGATGTTGAATATGCCGACTTTGTTGACGGAGATGGCTGAAGGGATGTTGATGA
GTCCACCGAGAATGATGATACATCCGACGATGGAAGATGATTCCGCCGAGAATCATGAAG
GAGATAATCTTTGGAGTTATAATGAATCCATTGAAGCTGCTCTTTTTTTATTGTTTTTC
CGGTGCAATGAGATTTTCCCCCTTTTTTTTTTCTTTTTGGGTGCGTGT

>G913 Amino Acid Sequence (domain in AA coordinates: 62-128)
MSNNNSPTTVNQETTSREVSITLPTDQSPQTSPPSSSPSPRPSGGSPARRTATGLSG
KHSIFRGIRLRNGKWVSEIREPRKTTRIWLGTYPVPEMAAAAYDVAALALKGPDAVLNFP
GLALTYVAPVNSAADIRAAASRAEMKQPDQGGDEKVLEPVQPGKEEELEEVSNCNSL
EFMDBEAMLNMPILLTEMAEGMLMSPPRMMIHPTMEDDSPENHEGDNLSYK*

>G937 (45..1046)

TGGAAAAAGTTTGAGTTTTTAATTCGAATCGAGAAAAATAAAATGGGTCTTTAGGTG
ATGAGCTTAGTTTGGGATCGATCTTTGGGAGAGGAGTTTCGATGAATGTTGTGGCGTTG
AGAAAGTTGATGAACATGTTAAGAAGCTTGAAGAAGAGAAGAGAAAGCTCGAAAGTTGTC
AACTTGAGCTTCTCTGTCTTTGTCAGATTTTAAACGATGCGATTTTGTATCTGAAGGATA
AGAGATGTTTCAGAGATGGAGACTCAACCATTTGTTGAAAGATTTCAATTTCTGTTAATAAAC
CTATTCAAGGAGAAAGAGGAATAGAATTGCTGAAAAGAGAGGAGCTAATGAGGGAGAAGA
AGTTTCAGCAATGGAAGCTAATGATGATCAGCTAGTAAGATCAAGAGCAAGCTTGAGA
TTAAGAGAAATGAGGAGAAATCTCTATGTTGTTGATTCCAAAGGTGGAACTGGTTTAG
GCCTCGGTTTAAAGTTTCGAGTTGATAAGAAGAAAAGGGATTGTTGCCTCATGTGGCTTTA
CTTCTAACTCTATGCCACAACCACCAACACCAGCAGTACCACAACAACCAGCATTCTTA

AGCAGCAAGCTTTACGGAAGCAAAGAAGGTGTTGGAATCCAGAGTTGCATCGCCGATTTG
TCGATGCATTGCAACAGCTAGGTGGACCGGGAGTGGCAACTCCTAAACAAATTAGAGAAC
ATATGCAAGAAGAAGGCTTAACCAATGATGAAGTCAAGAGTCATTTACAGAAATACAGGT
TACACATCAGGAAGCCAAATTCGAATGCGGAGAAACAATCAGCAGTTGTTTTAGGGTTTA
ACTTGTGGAATTCTTCAGCACAAGATGAAGAAGAGACATGTGAAGGAGGAGAATCATTGA
AGAGAAGCAATGCGCAATCAGATTCTCCTCAAGGTCCTTTGCAGTTACCGTCTACAACAA
CAACAACCTGGTGGAGATAGTAGCATGGAAGATGTTGAAGATGCTAAGTCTGAGAGCTTTC
AACTGGAGAGATTGAGATCACCATAAAATCTCAAGAAACCAAACTCTTGATCACGGTTTTG
TTATTTTGGATTCTATTACTATATCTATTAGTAGTGAATGAGAACAATAATTATAGAAAGG
TTTATAGATATATATATAGAGAAAAAGAGAGAGTGAGGATGTTCAAATTATTTGCAGA
>G937 Amino Acid Sequence (conserved domain in AA coordinates:197-246)
MGS LGDELSLGSIFGRGVMNVVAVEKVDEHVKKLEEEKRKLESCQLEPLSLQILNDAI
LYLKD KRCSEMETQPLLKDFISVKNPIQGERGIELLKREELMREKKFQQWKANDDHTSKI
KSKLEIKRNEEKSPMLLIPKVETGLGLSSSSIRRKGIVASCFTS NSMPQPPTPAVPQ
QPAFLKQQA LRKQRRWCNPELHRRFVDALQQLGGPGVATPKQIREHMQEEGLTNDDEVKSH
LQKYRLHIRKPNSNAEQSAVVLGFNLWNSSAQDEEETCEGGESLKRNSNAQSDSPQGPLQ
LPSTTTTTGDSMEDVEDAKSESFQLERLRSP*
>G960 (63..1538)
TACCGTCGACCCACGCGTCCGAGTGATTCAAAGTCGGAAAGAAACCTAAAGAAGAGGA
TTATGGGTGCTGTATCGATGGAGTCGCTTCCTTTAGGTTTCAGATTACAGACCTACCGATG
AAGAGCTCGTCAATCACTACCTCCGTCTCAAGATCAACGGACGTCACCTCCGATGTCCGTG
TCATCCCTGATATCGATGTCTGCAAATGGGAACCTTGGGATCTTCTGCTCTCTCGGTGA
TTAAGACGGATGATCCAGAGTGGTTCTTTTTCTGCCCTCGTGATCGGAAATACCCTAATG
GTCATCGCTCTAACAGAGCAACTGACTCTGGCTATTGGAAAGCTACTGGTAAAGATCGTA
GCATCAAGTCTAAGAAGACTTTAATCGGTATGAAGAAGACTCTTGTCTTCTATCGTGGAC
GAGCTCCTAAAGGTGAGCGGACTAATTGGATTATGCACGAGTATCGTCCCCTCTTAAGG
ATCTTGATGGCACTTCCCCTGGCCAAAGCCCTTACGTTCTTTGTGCGCTCTTCCACAAGC
CTGATGATCGGGTTAATGGTGTCAAGTCCGATGAAGCAGCTTTTACGGCCAGCAACAAAT
ACTCACCTGATGATACATCATCTGATCTTGTTCAGAAACACCTTCTCTGATGCTGCTG
TTGAGAAACCATCAGATTATTCAGGTGGATGCGGTTATGCTCATAGTAATAGTACCGCAG
ATGGGACAAATGATTGAGGCACCTGAAGAGAATCTTTGGTTATCTTGTGACCTTGAAGATC
AAAAGGCACCACTACCGTGTATGGATTCTATATATGCTGGTGATTTCAGTTACGATGAGA
TTGGATTCCAATTTCAAGATGGTACCAGCGAACCAGATGTATCACTAACAGAATTGTTGG
AGGAGGTGTTCAATAACCTGATGACTTCTCTTGGCAGGAATCGATCAGTCGAGAGAATC
CAGCAGTCTCACCAATGGGATATTTTCTCTGCTAAATGCTGCAGTCTGCAGCACCAG
AGGATGCTTTCTTCAACGACTTCATGGCTTTCAGTATACAGATGCTGAGATGGCGCAAT
TGCAGTATGGTTTCAAGAGTGGAGCTTCTGGTTGGCCAAGTGACACTAATTCATACTATA
GTGATTTGGTTTCAAGAGCAAATGATCAATCATAACACAGAGAACAACCTCACAGAAG
GGAGAGGGATAAAGATCCGGGCTCGACAGCCTCAGAACCAGCAGAGTACAGGATTGATAA
ACCAGGGTATTGCTCCAAGGAGAATCCGTCTGCAGCTGCAGTCTAATCTGAAGTAAAG
AACGAGAGGAGGTGAATGAAGGACACACTGTTATTCCCGAGGCCAAAGAAGCTGCAGCTA
AATACTCAGAGAAGAGTGGTTCTTTGGTTAAACCTCAAATAAAGCTCAGGGCGCGGGGAA
CTATAGGCCAAGTAAAGGAGAGAGATTTCAGACGACGAGGTACAGGTGCAGAGCACAA
AGAGAGAGAGAGAGAGAATCAAATGTAGTTTAATGTAATTAGGGATGATGCAATGTTAGC
ATGTTTGTGTGTTGTAACCTTAAAACTTATTAGGAATCTGATAAAAGTTACTGTTGAAA
AAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
>G960 Amino Acid Sequence (domain in AA coordinates: 13-156)
MGA VSMESLPLGFRFRPTDEELVNHYLRLLKINGRHSDVRVIPDIDVCKWEPWDLPALSVI
KTDDPEWFFFCPRDRKYPNGHRSNRATDSGYWKATGKDRS IKS KTLIGMKTLVFYRGR
APKGERTNWMHEYRPTLKLDGTSPGQSPYVLCRLFKHPDDRNVGVKSDEAAPTASNKY
SPDDTSSDLVQETPSSDAAVEKPSDYSGGCGYAHNSNADGTMIEAPEENLWLSCDLEDQ
KAPLPCMDSIYAGDFS YDEIGFQFQDGTSEPDVSLTELEEVFNPNDDFSCBESISREN
AVSPNGIFSSAKMLQSAAPEDAFFNDFMAFTDTDAEMAQLQYGSSEGGASGWPSTNYSYS
DLVQQEQMINHNNTENLTEGRGIKIRARQPQNRQSTGLINQGIAPRRIRLQLQSNSEVKE
REEVNEGHTVPIPEAKEAAKYSEKSGSLVKPQIKLRARGTIGQVKGERFADDEVQVQSTK
RERERIKCSLM*
>G991 (6..533)

GAAAAATGGAAGAAGAAAAGAGATTGGAGCTAAGGCTAGCTCCTCCTTGTACCAATTCA
CTTCCAACAACAACATCAATGGATCTAAACAAAAAAGCTCGACCAAAGAAACATCATTCC
TTTCCAATAACAGGGTTGAGGTAGCTCCAGTGGTGGGATGGCCGCCGGTGAGATCATCCC
GGAGAAACCTAACGGCACAACCTAAAGGAGGAGATGAAGAAGAAGGAGAGTGATGAAGAGA
AGGAATTGTACGTTTAAGATCAACATGGAAGGAGTTCCAATAGGAAGAAAAGTCAACCTTT
CAGCTTATAACAACCTACCAACAGCTTTTACATGCCGTTGACCAACTCTTCTCTAAGAAAG
ATTCTGGGATCTAAACAGACAATACACTTTGGTCTACGAAGACACTGAAGGAGATAAAG
TTCTGGTGGGGATGTTCTTTGGGAGATGTTTGTATCTACTGTAAAGAGGTTGCATGTTT
TAAAGACCTCCACGCCCTTCTCACTCTCACCTAGAAAAACATGGCAAGGAATAGAGAGAGG
TTGGCCAAAATCATCAGTTTCGATGGTTTGTCTTTAATGTAATTTTGTGGAACTAATGG
GGTTTGGCTTTGATTTACTGGTTTCTTTTCACTTATGTACTAGGTTTGTGCTTGCTAT
GTTATTTCTTGTGTTTGGTTTGAATATGCTGTTTAAAGAAATCGGGGGTTAGTATGT
TATCGTGTGTATAAAAATAGTGTAAAGCACGTAAGTTGATTACAAAAAATAAAAAA
AAAAA

>G991 Amino Acid Sequence (domain in AA coordinates: 7-14,48-59,82-115,128-164)

MEEEKRLELRAPPCHQFTSNNNINGSKQKSSTKETSFLSNRVEVAPVVGWPPVRSSRR
NLTAQLKEEMKKESDEEKELYVKINMEGVPIGRKVNLSAYNNYQQLSHAVDQLFSKKDS
WDLNRQYTLVYEDTEGDKVLVGDPWEMFVSTVKRLHVLKTSFAFSLSPRKHGKE*

>G748 (98..1444)

CCACGCGTCCGCACTCTCCCAAATCTCTCTTCTTTAACAACAAAAAATCACAGAGA
CATAGAGAGAAGAAGACGGAACAGAGGCTCCAAAAAATGATGATGGAGACTAGAGATCC
AGCTATTAAGCTTTTCGGTATGAAAATCCCTTTTCCGTCGGTTTTTGAATCGGCAGTTAC
GGTGGAGGATGACGAAGAAGATGACTGGAGCGGCGGAGATGACAAATCACCAGAGAAGGT
AACTCCAGAGTTATCAGATAAGAACAACAACACTGTAACGACAACAGTTTTAACAATTC
GAAACCCGAAACCTTGGACAAGAGGAAGCGACATCAACTGATCAGATAGAGAGTAGTGA
CAGCGCTGAGGATAATCAGCAGACGACACCTGATGGTAAACCCCTAAAGAAACCGACTAA
GATTCTACCGTGTCCGAGATGCAAAAGCATGGAGACCAAGTTCTGTTATTACAACAATA
CAACATAAACAGCCTCGTCATTTCTGCAAGGCTTGTGAGAGATATTGGACTGCTGGAGG
GACTATGAGGAATGTTCTGTGGGGGCAGGACGTCGTAAGAACAAGCTCATCTTCTCA
TTACCGTCACATCACTATTTCCGAGGCTCTTGAGGCTGCGAGGCTTGACCCGGGCTTACA
GGCAACACAAGGGTCTTGAGTTTGGTCTCGAAGCTCAGCAGCAGCACGTTGCTGCTCC
CATGACACCTGTTATGAAGCTACAAGAAGATCAAAAGGCTCTCAAACGGTGCTAGGAACAG
GTTTCACGGGTTAGCGGATCAACGGCTTGTAGCTCGGGTAGAGAATGGAGATGATTGCTC
AAGCGGATCCTCTGTGACCACCTTAACAATCACTCAGTGGATGAATCAAGAGCACAAG
CGGCAGTGTGTTGAAGCACAATGAACAACAACAACAATAACATGAATGTTATGC
TTGCATCCAGGTGTTCCATGGCCTTACACGTGGAATCCAGCGATGCCTCCACCAGGTTT
TTACCCGCTCCAGGTATCCAATGCCGTTTTACCCCTTACTGGACCATCCCAATGCTACC
ACCGCATCAATCCTCATCGCTATAAGCCAAAAGTGTTCAAATACAACTCTCCGACTCT
CGGAAAGCATCCGAGAGATGAAGGATCATCGAAAAGGACAATGAGACAGAGCGAAAACA
GAAGGCCGGGTGCGTTCTGGTCCCGAAAACGTTGAGAATAGATGATCCTAACGAAGCAGC
AAAGAGCTCGATATGGACAACATTGGGAATCAAGAACGAGGCGATGTGCAAGCCGGTGG
TATGTTCAAAGGGTTTGATCATAAGACAAAGATGTATAACAACGACAAAGCTGAGAACTC
CCCTGTTCTTCTGCTAACCTGCTGCTCTATCAAGATCACACAATTTCCATGAACAGAT
TTAGAGTTACATATGTATATGTATATGTATGATTGATTGTATGTATAGATGATACTGG
AGAATGATGAGTTTTTGAGAATCAAACCTCTTTCTTTCTTTCTAGTGATTGCCTTTATTCC
TTTACATGTTTGGTTCTCTGTACACTATTTGATTACCTTTTTTACTTTCTTTCTTCAT
TTGTCAGGAAATGTTGGAAGATAACATTAATGGTAAAAAGTTGGTGTGGACCGTTGTTGC
GTTGGCATTTCAAAAAATAAAAAA

>G748 Amino Acid Sequence (domain in AA coordinates: 112-140)

MMMETRDPAILKFGMKIPFVSFESAVTVEDEEDDWSGGDDKSPEKVTPELSDKNNNNC
NDNSFNNSKPELTLDKEEATSTDQIESSTPEDNQQTTPDGKTLKKPTKILPCPRCKSMET
KFCYYNNYINQPRHFCACQRYWTAGGTMNRNVPVAGRRKNKSSSSHYRHITISEALEA
ARLDPLQLANTRVLSFGLEAQQHVAAAPMTPVMKLQEDQKVSNGARNRFGHLADQRLVAR
VENDDCSSGSSVTSNNHSVDESRAQSGSVVEAQMNNNNNNNMNGYACIPGVWPYPYTN
PAMPPPGFYPPGYPMPFYPYWTIPMLPPHQSSSPISQKCSNTNSPTLGKHPRDEGSSKK
DNETERKQKAGCVLPKTLRIDDPNEAAKSSIWTTLGIKNEAMCKAGGMFKGFDHKTMY
NNDKAENSPVLSANPAALSRSHNFHEQI*

>G247 (1..660)
ATGAGAATGACAAGAGATGGAAAAGAACATGAATACAAGAAAGGTTTATGGACAGTGGAA
GAAGACAAGATCCTCATGGATTATGTCCGAACATCATGGCCAGGGCCACTGGAACCGCATC
GCCAAGAAAACCTGGGCTCAAGAGATGTGGGAAAAGCTGTAGGTTGAGATGGATGAACAC
TTAAGCCCCTAATGTTAACAGAGGCAATTTTACTGACCAAGAAGAAGATCTCATCATCAGA
CTCCACAAGCTCCTCGGCAACAGATGGTCGTTGATAGCGAAAAGAGTTCCGGGAAGAACA
GACAACCAAGTAAAGAAATTACTGGAACACACATCTCAGCAAGAACTTGGTCTCGGAGAT
CATTCAACTGCCGTCAAAGCCGCATGCGGTGTAGAGTCTCCACCGTCTATGGCCCTTATA
ACCACAACGTCCTCCTCATCAAGAGATCTCCGGTGGAAAAAATCAACTCTAAGGTTTC
GACACTTTTAGTTGACGAATCCAACTCAAACCAAAATCCAACTAGTCCACGCAACACCA
ACTGACGTAGAAGTTGCAGCTACGGTTCCAAATCTGTTCGATACCTTTTGGGTTCTTGAA
GACGACTTCGAGCTTAGTTCACTCACTATGATGGATTTTACTAATGGGTATTGCTTTTGA
>G247 Amino Acid Sequence (domain in AA coordinates: 15-116)
MRMTRDQKEHEYKKGLWTVVEEDKILMDYVRTHGQGHWNRIAKKTGLKRCGKSCRLRWMNY
LSPNVNVRGNFTDQEEELIIRLHKLLGNRWSLIAKRVPGRDNDQVKNYWNTHLSKKLGLGD
HSTAVKAACGVESPPSMALITTTSSSHQEIISGGKNSTLRFDTLVDESKLKP KSKLVHATP
TDVEVAATVFNLFDTFWVLEDDFELSSLTMMDFTNGYCL*
>G585 (111..2039)
CTCTCAAACATTTCTCTGTTTGTTCGGCGAAAACGGCAACTGTTTCATCAAATGACAAA
CACAAAACCTTAACATCTAGTTTGTATCCTCTCTGATACCTCAAAAAAATGGATGAAG
AAACAATGGCTACCGGACAAAACAGAACAACTGTGCCAGAGAATCTGAAGAAACACCTCG
CAGTTTCAGTTCGAACATTCAATGGAGTTATGGTATCTTTTGGTCTGTCTCTGCTTCTC
AGTCTGGAGTTTTAGAATGGGGAGATGGATACTATAATGGAGATATCAAACGAGGAAGA
CGATTCAAGCTTCGGAGATCAAAGCTGATCAGCTTGGTCTACGGAGGAGCGAGCAGCTTA
GCGAGCTTTACGAGTCTCTCTCCGTCGCTGAATCTTCTTCTCAGGCGTTGCTGCCGGAT
CTCAAGTCACCAGACGAGCTTCCGCCGCCGCACCTTTCACCGGAAGATCTCGCCGACACCG
AGTGGTACTATTTTGGTTTGTATGTTCTTCTGTTCTCAACATTGGTGAAGGAATGCCTGGAC
GGACGTTTGCAAACGGTGAACCGATATGGTTGTGCAACGCTCATACGGCGGATAGTAAAG
TGTTTTAGCCGTTCTCTTCTAGCAAAAAGTGCTGCGGTTAAGACAGTGGTTTGCTTCCCGT
TCCTTGGAGGAGTCGTTGAGATTGGTACCACAGAACATATTACGGAAGACATGAATGTAA
TACAATGCGTGAAGACATCATCTCCTCGAAGCCCTGATCCGTACGCTACAATATTACCAG
CAAGATCCGATTATCACATCGACAACGTTCTTGATCCGCAACAGATTCTAGGCGACGAGA
TTTACGCGCCTATGTTCACTACGAGCCTTTTCCAACAGCTTCTCCGAGCAGAACTACCA
ACGGTTTTCGATCAAGAACATGAACAAGTAGCAGATGATCATGATTCTTTTCATGACCGAAA
GAATCACTGGAGGAGCTTCTCAGGTGCAAAGCTGGCAGCTCATGGACGACGAGCTTAGTA
ACTGCGTTTACCAGTCTGCTAAATTCCAGCGATTGCGTCTCTCAAACGTTTGTGTAAGGGG
CGGCTGGACGGGTTGCTTACGTTGCAAGAAAGAGTAGAGTTCAAAGACTAGGGCAAATTC
AAGAGCAACAGAGAAATGTGAAGACATTGTCAATTTGATCCAAGAAACGACGACGTTTCATT
ACCAAAGTGTGATCTCAACGATTTTAAAGACCAACCATCAGTTAATTCTCGACCGCAGT
TTCGAAACTGCGATAAACAGTCAAGCTTCACTAGGTGGAAGAAATCATCGTCATCATCAT
CAGGAACCGCCACGGTCACGGCACCATCAAGGAATGTTAAAGAAAATTATTTTCGATG
TTCCGCGAGTGACACCAGAAAGAGAAGTTAATGTTGGACTCACAGAAAGCCAGAGATGAAA
CTGGGAACCATGCGGTTTTAGAGAAGAAGCGCCGAGAAATTGAACGAACGGTTCATGA
CCTTGAGAAAAATCATTCCGTCAATCAACAAGATCGATAAAGTATCGATTCTTGACGATA
CGATAGAGTATCTTCAAGAACTCGAGAGACGGGTTCAAGAACTAGAACTTTCGAGAGAAT
CAACCGATACAGAGACTCGTGGGACGATGACGATGAAGAGGAAGAAACCATGCGACGACG
GAGAAAGAACATCAGCTAATTGCGCAAATAATGAAACAGGAAATGGGAAGAAGGTGTCGG
TTAACAATGTTGGTGAAGCCGAGCCAGCAGATACCGGTTTACTGGTTTAAACCGATAATT
TAAGGATCGGTTTCGTTTGGTAATGAGGTGGTTATTGAGCTTAGATGTGCTTGGAGAGAAG
GAGTATTGCTTGAAGATAATGGATGTGATTAGTGATCTCCATTTGGATTCTCATTCCGGTTC
AATCCTCGACCGGAGACGGTTTGTCTGTCTTAACCGTCAATTGCAAGCACAAGGGGTCAA
AAATAGCGACACCAGGAATGATCAAAGAAGCACTTCAAAGGGTTGCATGGATCTGTTGAA
GACTACTTAGTTAAATTTGACAGCAAAGAAAAACATTCCCGGTTTGGTTTTCTATTCTTT
GGTTTTCTTCTAACCGGTTTTAGGAATTAATGTTATGTTTATCATTTGTTTTTGTGTTTT
TTTTTTGTGTCTTTTTTTCCGTTGCTTAACGTAGGTGAAGAGGAACATACTATGCGTA
TTTTGTTTGGGTAGATTATTTAAGGGTATTAGTAATAGTAATAGCCAGTTTAGATGAT
TTTGTGTTCTTTTGTGTT

>G585 Amino Acid Sequence (domain in AA coordinates: 436-501)
MDEETMATGQNRTTVPENLKKHLAVSVRNIIQWSYGIFWSVSASQSGVLEWGDGYNGDIK
TRKTIQASEIKADQLGLRRSEQLSELYESLSVAESSSSGVAAGSQVTRRASAAALSPEDL
ADTEWYYLVCMSFVFNIGEGMPGRTFANGAPIWLCNAHTADSKVFSRSLAKSAAVKTVV
CFPFLGGVVEIGTTEHITEDMNVIQCVKTSFLEAPDPYATILPARSDYHIDNVLDPPQIL
GDEIYAPMFSTEPFPTASPSRTTNGFDQEHEQVADDHDSFMTERITGGASQVQSWQLMDD
ELSNCHVQSLNSSDCVSQTFVEGAAGRVAYGARKSRVQRLGQIQEQQRNVKTLSDPRND
DVHYQSVISTIFKTNHQLILGPQFRNCDKQSSFTRWKKSSSSSSGTATVTAPSQGMKKKI
IFDVPRVHQEKLMMLDSPEARDETCNHAVLEKKRREKLNRFMTLRKIIPSINKIDKVS
LDDTIEYLQELERRVQLESCRESTDTETRGTMKRRKKPCDAGERTSANCANNETGNGK
KVSNNVGEAEPADTGFTGLTDNLRIGSFGNEVIELRCAWREGVILIMDVISDLHLDS
HSVQSSTGDGLLCLTVNCKHKGSKIATPGMIKEALQRVAWIC*

>G634 (1..798)

ATGGAGCAAGGAGGAGGTGGTGGTGGTAATGAAGTTGTGGAGGAAGCTTCACCTATTAGT
TCAAGACCTCCTGCTAACAACTTAGAAGAGCTTATGAGATTCTCAGCCGCCGCGGATGAC
GGTGGATTAGGAGGTGGAGGTGGAGGAGGAGGAGGAAGTGCTTCTTCTCATCGGGA
AATCGATGGCCGAGAGAAGAACTTTAGCTCTTCTTCGGATCCGATCCGATATGGATTCT
ACTTTTCGTGATGCTACTCTCAAAGCTCCTCTTTGGGAACATGTTTCCAGGAAGCTATTG
GAGTTAGGTTACAAACGAAGTTCAAAGAAATGCAAAGAGAAATCGAAAACGTTCAAGAA
TATTACAAACGTACTAAAGAACTCGCGGTGGTCGTCATGATGGTAAAGCTTACAAGTTC
TTCTCTCAGCTTGAAGCTCTCAACACTACTCCTCCTCCTCCTCCTTCTCATCCTCACGCT
CATCAACCAGAACAGAAACAACAACAACCAACCAAGAGATGGTCATGAGCTCGGAA
CAATCATCATTACCATCATCATCAAGATGGCCAAAGGCAGAGATTCTAGCGCTTATAAAC
CTGAGAAGTGAATGGAACCAAGGTACCAAGATAATGTACCTAAAGGACTTCTATGGGAA
GAGATCTCAACTTCAATGAAGAGAATGGGATACAACAGAAACGCTAAGAGATGTAAAGAG
AAATGGGAAAACATAAAACAATACTACAAGAAAGTTAAAGAAAGCAACACAGCAACTAC
AACAAACAAGATCAATGA

>G634 Amino Acid Sequence (domain in aa coordinates: 62-147, 189-245)

MEQGGGGGGNEVVEEASPISSRPANNLEELMRFSAAADDGGLGGGGGGGGGSASSSSG
NRWPREETLALLRIRSDMDSTFRDATLKAPLWEHVSRLLELGYKRSSKKCKEKFENVQK
YYKRTKETRGRHDGKAYKFFSQLEALNTTPPPPPSHPHAHQPEQKQQQPQOEMVMSSE
QSSLPSSSRWPKAELALINLRSGMEPRYQDNVPGLLWEEISTSMKRMGYNRNAKRCKE
KWENINKYKVKESNNSNYNNKNQ*

>G676 (1..612)

atgagaaagaaagtaagtagtagtggtgacgaaggaaacaatgagtacaagaaagggttg
tggaacagtagaagaagacaaaatcctcatggattatgtcaaagctcatggcaaagggtcac
tggaatcgtattgccccaaaagactgggttaaaagagatgtggaaagagttgtagattgagg
tggtatgaattatctcagccctaattgtgaaaagaggcaatttcaccgagcaagaaggat
cttatcattaggtccacaagttgcttggtaataggtggtctttaattgtcaaaagagt
ccgggtcgaaacggataaatcaagtgagaactattggaacacgcacatcttagtaagaaactc
ggaatcaaagatcagaaaacaaacagagcaatgggtgatattgtttatcaaatcaatctc
ccgaatcctaccgaacatcagaagaaacgaaaatctcgaatattgtcgataacaataat
atcctcggagatgaaattcaagaagatcatcaaggaagtaactacttgagttcactttgg
gttcacgagtaggattgagcttagcacactcaccaacatgatggactttatagatgga
cactgtttttga

>G676 Amino Acid Sequence (domain in AA coordinates: 17-119)

MRKVVSSSGDEGNNEYKGLWTVEEDKILMDYVKAHGKGHWNRIAKKTGLKRCGKSCRLR
WMNYLSPNVKRGNFTEQEEDLIIRLHKLLGNRWSLIAKRVPGRTDNQVKYWNTHLSKKL
GIKDKTKQSNQDIVYQINLPNPTETSEETKISNIVDNNILGDEIQEDHQGSNYLSSLW
VHEDEFELSTLTNMMDFIDGHCF*

>G682 (1..228)

ATGGATAACCATCGCAGGACTAAGCAACCCAAGACCAACTCCATCGTTACTTCTTCTTCT
GAAGAAGTGAGTAGTCTTGAGTGGGAAGTTGTGAACATGAGTCAAGAAGAAGAAGATTG
GTCTCTCGAATGCATAAGCTTGTCTGGTGACAGGTGGGAAGTATAGCTGGGAGGATCCCA
GGAAGAACCCTGGAGAAATTGAGAGGTTTTGGGTCATGAAAAATTGA

>G682 Amino Acid Sequence (domain in AA coordinates: 27-63)

MDNHRRTKQPKTNSIVTSSEEVSSLEWEVVMMSQEEDLVSRMHKLVGDRWELIAGRIP

GRTAGEIERFWVMKN*

>G635 (1..993)

ATGGAGATCATGCGTCCAGGGGTCTCAGAAAACACTTTGAAAGGAAAAATAAGAATCACA
ACGCGGTGCATGTGGCTTGACAAAGGAAGACTTTTAGATGCACCTTCAAAAGCAGCTCAT
GCTGCTCTATCAAGTTGTCTGTGACATGTCCCTGTCTCACATGGAAAGAACAGTCTCC
GAAGTCTTGAGGAAGATTGTAAGGAAGTACAGTGGTAAAGGCCTGAAGTCATCGCTATA
GCCACTGAGAATCCAAATGGCTGTCCGAGCTGATGAGGTCAAGTGCAGAGACTGTCTGGTGAT
CCAAGTGTGGTTCTGGAGTTGCAAGCTTTAAGGAAAGTTGTTGAAGGAAATGACAAAAGA
AGTCGGGCGAAGAAAGCACCTTCACAAGAAGCTTCCCCCAAAGTAGATCGCACTTTG
GAAGATGATATCATTGATAGTGCAAGACTACTGGCTGAAGAAGAACTGCGGCATCAACA
TACACGGAAGAAGTTGATACGCCCGTTGGGAGTTCTTCAGAAGAGTCAGACGATTTTTGG
AAATCATTTCATCAATCCATCATCGTCACCTTCACCGAGTGAAACAGAAAATATGAATAAG
GTAGCTGATACGGAGCCTAAAGCAGAGGGTAAGGAAAACAGCAGAGACGACGATGAATTA
GCTGATGCTTCAGATTCTGAAACCAAGTCATCACCAAAACGTGTGAGGAAGAACAATGG
AAACCGGAGGAGATAAAGAAGGTAAATCAGAATGCGAGGAGAGCTGCACAGTAGATTTCAA
GTGGTGAAAGGTAGAATGGCATTGTGGGAAGAGATCTCTTCAAATCTATCAGCTGAAGGA
ATCAATCGAAGCCCGGACAATGCAAATCTCTCTGGGCATCACTTATTTCAGAAATACGAG
GAGAGCAAGGCTGATGAGAGAAGCAAGACGAGTTGGCCACATTTTGAGGATATGAACAAC
ATTTTGTCTAGAGCTAGGCACACCTGCGTCTTAA

>G635 Amino Acid Sequence (domain in AA coordinates: 239-323)
MEIMRPGVSENTLK GKIRITTRCMWLDKGRLLDALHKAHAALSSCPVTCPLSHMERTVS
EVLRLKIVRKYSGRPEVIAIATENPMVRADSVSARLSGDPVSGSGVAALRKVVEGNDKR
SRAKKAPSQEASPKVDRTLEDDIIDSARLLAEETAASTYTEEVDTPVGSSEESDDFW
KSFINPSSSPSPSETENMNKVADTEPKAEGKENSRRDDELADASDSEKSSPKRVIRKNKW
KPEEIKKVIRMRGELHSRFQVVKGRMALWEEISSNLSAEGINRSPGQCKSLWASLIQKYE
ESKADERSKTSWPHFEDMNNILSELGTPAS*

>G1068 (150..1310)

GAGAGTTGTTAGCTAGCTCACACGCTTTTCGCTTAAACTCAAAACCTGCACCTTTCTCGT
CTATTTTCTCGGCATTTCGTAAGAACAGAAAAGTGGGTCTCCAAGAAAATTACCCTAAATTC
ACAAAGATTCTACTTTTCTCCACCTCCAATGGATTCCAGAGAGATCCACCACCAACAAC
AGCAACAACAACAACAACAGCAGCAGCAGCAACAACAGCAACATCTACAACAACAGC
AACAACCACCGCCAGGGATGTTAATGAGTCACCACAATTCCTACAATCGAAACCCCTAACG
CCGCCGCCGCTGTTTAAATGGGTCAACAACCTCCACATCTCAAGCTATGCATCAAAGAT
TACCTTTTGGTGGTTCTATGTCACCGCATCAGCCTCAACAACATCAGTATCATCATCCTC
AGCCTCAGCAACAGATAGATCAGAAGACTCTTGAATCTCTTGGATTTCCTACTTCGCCTC
TTCCTTCTGCTTCTAATTTCTTACGGTGGTGGAAATGAAGGAGGTGGTGGTGGTAGCG
CCGAGCTAATGCTAATCTTCCGATCCACCTGCTAAACGGAACAGAGGACGCTCCTCTG
GCTCCGGTAAGAAGCAGCTCGATGCTTTAGGAGGAACAGGAGGAGTTGGGTTACGCCTC
ATGTCATTGAGGTTAAACAGGAGAGGACATAGCTACGAAGATATTGGCGTTTACGAACC
AAGGCCACGCGCAATCTGTATTCTCTCAGCTACAGGAGCTGTAACATAATGTGATGCTTC
GTCAAGCTAACAATAGCAATCCTACTGGAAGTGTAAAGTATGAGGGCCGATTTGAAATCA
TTTCTCTGTCAAGTTCTTTCTTGAATTTCTGAGAGTAATGGTACTGTGACCAAACTGGTA
ACTTGAGTGTGTCGCTGGCTGGACACGAAGGCCGATTGTGGGTGGATGTGTTGATGGAA
TGCTAGTAGCTGGATCACAAGTCCAGGTCATTGTGGGAAGCTTTGTACCAGATGGAAGGA
AGCAGAAAACAAAGTCCGGGGCGTGCTCAGAATACTCCGGAGCCAGCTTCAGCACCCAGCCA
ATATGTTGAGCTTTGGTGGTGGTGGTGGACCGGAAGCCCTCGATCTCAAGGACAACAAC
ACTCGAGCGAGTCATCAGAGGAAAACGAAAGTAATTTCTCCGTTGCACCGTAGAAGCAACA
ACAACAACAGCAACAATCATGGGATATTGGAACTCTACACCTCAACCGCTTCACCAAA
TTCCTATGCAGATGTACCAGAATCTCTGGCCTGGCAACAGTCTCAATAAACAGATGGTT
CATGGGTCAAGATTTGACCGGGTTTGCTTCTCTGTTCTTTTGACACATCTCTCCATCAG
ATTTATCTCTATAAAGTAGATTGAGCTCTCTTACTCTCTCATCTTCTTCTCCTTTACTAT
TTCTCTTAAATTTAGCTTTGGTTTGTAGATAAATAGAGAGAGAGACATGTTAAGTAGGT
TTCAAATTCATCTTGTGTTAGTTGTTTCTTAGTAGTTTCTTTGATTGTGATGATCATA
AAGACTTGTTCTTTTCTCCTATATTCAACGAATTATCCACTTTAA

>G1068 Amino Acid Sequence (domain in AA coordinates: 143-150)
MDSREIHHQQQQQQQQQQQQQQQHLQQQQQPPPGMLMSHHNSYNRNPAAAAVLMGHN
TSTSQAMHQRLPFGGSMSPHQPHQYHHPQPPQIDQKTLES LGFPTSPLPSASNSYGG

GNEGGGGGDSAGANANSSDPPAKRNRGRPPGSGKKQLDALGGTGGVGFTPHVIEVKTGED
IATKILAFITNQGPRAICILSATGAVTNVMLRQANNNSNPTGTVKYEGRFEIISLSGSFLNS
ESNGTVTKTGNLSVSLAGHEGRIVGGCVDGMLVAGSQVQVIVGSFVPDGRKQKQSAGRAQ
NTPEPASAPANMLSFGGVGGPGSPRSQGGQHSSESSEENESNSPLHRRSNNNNSNNHGIF
GNSTPQPLHQIPMQMYQNLWPGNSPQ*

>G1225 (1..984)

ATGACTCTAGAAGCTTTATCATCAAACGGTCTTTTAAACTTTTGTCTCTGAAACTCTT
TCACCAACTCCATTCAGTCTCTCGTCGATCTCGAGCCATTGCCGGAATAATGATGTCTATC
ATATCGAAGAACAATTTTCGAGATATCTAATCAAGAACCACCACAGCGACAACCA
CCAGCTACGAATCGAGGGAAGAAGCGGCGGAGGAGGAAGCCTAGGGTTTGCAAAAACGAG
GAAGAAGCTGAGAATCAACGAATGACTCACATTGCCGTGAAAGAAATCGAAGAAGACAA
ATGAATCAACATCTCTCTGTCTTGCATCTCTCATGCCTCAACCTTTTGTCTACAAGGGT
GATCAAGCTTCAATAGTTGGTGGAGCCATAGATTTTCATCAAAGAATTGAACACAAATTA
CTATCTCTTGAAGCTCAAAAACATCATAATGCTAAATTAAACAGTCGGTTACTTCTTCA
ACAAGTCAAGACTCAAATGGTGAACAAGAGAACTCTCATCAACCATCTTCACTATCTCTA
TCGCAGTTCTTTCTTCAATTCATACGATCCGAGCCAAGAGAATAGGAACGGCTCAACAAGC
TCGGTGAAAACCCCTATGGAAGATCTTGAGGTGACTCTAATCGAACTCATGCTAACATC
AGAATCTTGTGCGAGAAGAAGAGTTTCCGGTGGAGCAGTTGGCCACCACCAACCGCCG
CAGCTTTCGAAGCTGGTGGCTTCTCTACAATCGTGTCCCTCTCCATTCTTCACTTAGT
GTCACAACTTGGACAATTATGCTATTTACTCCATCAGCGCTAAGGTGGAAGAGAGTTGC
CAGCTAAGTTCACTAGATGACATTGCAGGAGCAGTTCAACCATGCTAAGTATCATTGAA
GAGGAGCCTTTTGTGTCTCATCAATGTCAGAATTACCATTGACTTCTCTTTGAATCAC
TCAAATGCTCACTATCTCTCTGAGAAATCTCTTTTGTGTGTGTTATTCCTCTCTTTTA
ATTTTATCACATAGCACATCTTTAGTTTTTTTTTTTT

>G1225 Amino Acid Sequence (domain in AA coordinates: 78-147)

MTLEALSSNGLLNFLSETLSPTPFKSLVDLEPLPENDVIIISKNTISEISNQEPPPQRP
PATNRGKRRRRKPRVCKNEEEAENQRMTHIAVERNRRRQMNQHSVLRSLMPQPFHAKG
DQASIVGGAIDFIKELEHKLSSLEAQKHNAKLNQSVTSSTSDSQSNGEQENPHQPSLSL
SQFFLHSDPSQENRNGSTSSVKTPEMEDLEVTLIETHANIRILSRRRGFRWSTLATTKPP
QLSKLVASLQSLSLSLILHLSVTTLDNYAIYSISAKVEESCQLSSVDDIAGAVHMLSLIE
EFPFCCSSMSELPDFSLNHSNVTHSL*

>G1337 (97..1398)

AATGGATTTGTTCATCATCTTCTCACCCTCCTTAGTCTCTGAAAATAAATTCGATTTTG
ATTTTCGAATTTTAGGGATTTTGAGAGAGAGTCAGTTATGAGTAGTTTCGGAGAGAGTACCG
TGCGATTTCTGCGCGGAGCGTACGGCGGTTTTGTTTTGTAGAGCCGATACGGCGAAGCTG
TGTTTGCCTTGTGATCAGCAAGTTACACGGCGAATCTGTTGTCGAGGAAGCACGTGCGA
TCTCAGATCTCGGATAATTGCGGTAACGAGCCAGTCTCTGTTCCGGTGTTCACCGATAAT
CTGATTTTGTGTCAGGAGTGTGATTGGGATGTTACGGAAGTTGTTCACTTTCCGATGCT
CATGTTTCGATCCGCGTGAAGGTTTTTCCGGTTGTCCATCGGCGTTGGAGCTTGCTGTCT
TTATGGGGACTTGATTTGGAGCAAGGAGGAAAGATGAAGAGAATCAAGTTCCGATGATG
GCGATGATGATGGATAATTTCGGGATGCAGTTGGATTCTTGGGTTTTGGGATCTAATGAA
TTGATTGTTCCCGAGCATACGACGTTTAAAGAAGCGTGGATCTTGTGGATCTAGTTGTGGG
AGGTATAAGCAGGTATTGTGTAAGCAGCTTGAGGAGTTGCTTAAGAGTGGTGTGTCGGT
GGTGATGGCGATGATGGTGTGTCGACCGTGATTGTGACCGTGAGGGTGCTTGTGATGGA
GATGGAGATGGAGAAGCAGGAGAGGGGCTTATGGTTCCGGAGATGTCAGAGAGATTGAAA
TGGTCAAGAGATGTTGAGGAGATCAATGGTGGCGGAGGAGGAGGATTAAACGACAGTGG
AATGCTACTACTACTAATCTTAGTGGTGGCCAGAGTTCTCAGATATGGGATTTTAACTTG
GGACAGTCACGGGGACCTGAGGATACGAGTCGAGTGAAGCTGCATATGTAGGGAAAGGT
GCTGCTTCTTCATTCACAATCAACAATTTTGTGACCATATGAATGAACTTGTTCCACT
AATGTGAAAGGTGTCAAAGAGATTAAAAAGGATGACTACAAGCGATCAACTTCAGGCCAG
GTACAACCAACAAAATCTGAGAGCAACAATCGTCCAATTACCTTTGGCTCTGAGAAAGGT
TCGAACTCCTCCAGTGACTTGCAATTCACAGAGCATATTGCTGGAAC TAGTTGTAAGACC
ACAAGACTAGTTGCAACTAAGGCTGATCTGGAGCGGCTGGCTCAGAACAGAGGAGATGCA
ATGCAGCGTTACAAGGAAAAGAGGAAGACACGAGATATGATAAGACCATAAGGTATGAA
TCGAGGAAGGCAAGAGCTGACACTAGGTTGCGTGTGAGGCGAGATTGTGAAAGCTAGT
GAAGCTCCTTACCCTTAACCTTAAGTTTTTTTACATAGGCTTCTTTTAGCTACAACTT
AGTTACTTTTTTTTACTCCACTGCCTCATAAATGTACAGACCGGTCTCGTTTCATCTGGCC

GCCCTTCTTGTGTTTATTGCCTTATCTGGCCCTTTTATGTACCTTGAATCTTATCTAGTT
TAAAAAAGATTGTAACTTCTAGAAAACCATATTCTGTTGACAGTATATACATGCTCTATC
CAAGCAAAAA

>G1337 Amino Acid Sequence (domain in AA coordinates: 9-75)
MSSSERVPCDFCGERTAVLFCRADTAKLCLPCDQVHTANLLSRKHVRSQICDNCNEPV
SVRCFTDNLILCQECDDVHSGSCSVSDAHVRSVEGFGSCPSALELAALWGLDLEQGRKD
EENQVPMAMMDNFMQLDSWVLGSNELIVPSDTTFKKRGSCGSSCGRYKQVLCKQLEE
LLKSGVVGGDGDGDRDRDCDREGACDGDGDGEAGEGLMVPFEMSERLKWSDVEEINGGG
GGGVNQWNATTTNPSSGQSSQIWDNFNLGQSRGPEDTSRVEAAYVGKGAASSFTINNFD
HMNETCSTNVKGVKEIKDDYKRSTSGQVQPTKSESNNRPITFGSEKGSNSSDLHFTEH
IAGTSCKTTRLVATKADLERLAQNRGDAMQRYKEKRKTRRYDKTIRYESRKARADTRLRV
RGRFVKASEAPY*

>G1759 (110..700)
CGAGAAAAGGAAAAAATAGAAAGAGAAAACGCTTAGTATCTCCGGCGACTTGAAC
CCAAACCTGAGATCAAATTAGGGCACAAAGCCCTCTCGGAGAGAAGCCATGGGAAGAAA
AAAAGTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAAGTACCTTCTCCAAACG
TCGCAACGGTCTCATCGAGAAAGCTCGTCAGCTTTCTGTTCTCTGTGACGCATCCGTCTGC
TCTTCTCGTCTCTCCGCCCTCCGGCAAGCTCTACAGCTTCTCCTCCGGCGATAACCTGGT
CAAGATCCTTGATCGATATGGGAAACAGCATGCTGATGATCTTAAAGCCTTGGATCATCA
GTCAAAAGCTCTGAATATGGTTCACACTATGAGCTACTTGAACCTTGTGGATAGCAAGCT
TGTGGGATCAAATGCAAAAATGTGAGTATCGATGCTCTTGTTCACCTGGAGGAACACCT
TGAGACTGCCCTCTCCGTGACTAGAGCAAGAAGACCGAACTCATGTTGAAGCTTGTGA
GAATCTTAAAGAAAAGGAGAAAATGCTGAAAGAAGAGAACAGGTTTGGCTAGCCAGAT
GGAGAATAATCATCATGTGGGAGCAGAAGCTGAGATGGAGATGTCACCTGCTGGACAAAT
CTCCGACAACTCTCCGGTACTCTCCACTACTTAATTAGCCACCTTAAATCGGCGGTTG
AAATCAAAATCCAAAACATATATAATTATGAAGAAAAAATAAGATATGTAATTATT
CCGCTGATAAGGGCGAGCGTTTGTATATCTTAATACTCTCTCTTTGGCCAAGAGACTTTG
TGTGTGATACTTAAGTAGACGGAAGTCAATACTATCTGTTTAAAGACAAAAGGTTG
ATGAACCTTTGTACCTTATTCGTGTGAGAAAAA

>G1759 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)
MGRKKLEIKRIENKSSRQVTFSKRRNGLIEKARQLSVLCDASVALLVVSASGKLYSFSSG
DNLVKILDRYQKHADDLKALDHQSKALNYGSHYELLELVDSKLVGSNVKNVSIDALVQL
EEHLETALSVTRAKTEMLKLVENLKEKEKMLKEENQVLASQMNHHVGAEEAEMEMSP
AGQISDNLPVTLPLLN*

>G1804 (169..1497)
TATCTCTCTCTTTCTCAAAACCTTTTCAGTCAAAATTCTCCGGCGGCTTTTAAACTATGTG
AAGGAGGAGAACCTCCATAACAAGAAGCGGATTCTCTCAGTTTTCGGCGGCGGAGGAAC
ACAAAGCCACCGGTTTTAGACACACAGATTTTCAATTTTCAGTTGTTAAATGGTAAGTAGA
GAAACGAAGTTGACGTCAGAGCGAGAAGTAGAGTCCATGGCGCAAGCGAGACATAAT
GGAGGAGGTGGTGGTGAAGATCATCCGTTTACTTCTTTGGGAAGACAATCCTCTATCTAC
TCATTGACCCCTTGACGAGTTCCAACATGCTTTATGTGAGAACCAGCAAGAACTTTGGGTCC
ATGAACATGGACGAGTTTCTTGTCTCTATTTGGAACGAGAGGAGAATAATAACAATCAA
CAACAAGCAGCAGCAGCTGCAGGTTCACTTCTGTTCCGGCTAATCAAAATGGTTTCAAC
AACACAATAACAATGGAGGCGAGGGTGGTGTGGTGTCTTTAGTGGTGGTTCTAGAGGC
AACGAAGATGCTAACAAATAAGAGAGGGATAGCGAACGAGTCTAGTCTTCTCGACAAGGC
TCTTTGACACTTCCAGCTCCGCTTGTAGGAAGACTGTTGATGAGGTTTGGTCTGAGATA
CATAGAGGTGGTGGTAGCGGTAATGGAGGAGACGAAATGACGTAGTAGTAGTAAT
GGACAGAACAAATGCTCAGAACGGCGGTGAGACTGCGGCTAGACAACCGACTTTTGGAGAG
ATGACACTTGAGGATTTCTTGTGAAGGCTGGTGTGGTTAGAGAACATCCCACTAATCCT
AAACCTAATCCAAACCCGAACCAAAACCAAAACCGTCTAGTGTAATACCCGAGCTGCA
CAGCAACAGCTTTATGGTGTGTTTCAAGGAACCGGTGATCCTTCATTCGCGGTCAAGCT
ATGGGTGTGGGTGACCCATCAGGTTATGCTAAAAGGACAGGAGGAGGAGGTATCAGCAG
GCGCCACCAGTTTCAGGCAGGTGTTGCTATGGAGGTGGCGTTGGGTTTGGAGCGGTTGA
CAGCAAAATGGGAATGGTTGGACCGTTAAGCCCGGTGCTTCAGATGGATAGGACATGGA
CAAGTGGATAACAATAGGAGGTAGTATGGAGTAGATATGGGAGGGCTAAGGGGAAGGAAA
AGAGTAGTGGATGGTCCAGTGGAGAAAGTAGTGGAGAGAAGACAGAGGAGGATGATCAAG
AACCGCGAGTCTGCTGCTAGATCTAGAGCAAGAAAACAAGCATATACAGTGAATTGGAA

GCTGAACTTAACCAGTTGAAAGAAGAGAATGCGCAGCTAAAAACATGCATTGGCGGAGTTG
GAGAGGAAGAGGAAGCAACAGTATTTTGAGAGTTTGAAGTCAAGGGCACAAACCGAAATTG
CCGAAATCGAACGGGAGATTGCGGACATTGATGAGGAACCGAGTTGTCCACTCTAAACA
ACAATAGGAAGATGGAGAAGAAGTCGGAGACAGAACGAGGGAAAAACTGATGATTTTCT
ACGTTGTTGTTTGTCTTTTGAGGAATGAGGTTATAGAATCTTTATACTTTGATGTTTCT
GTGTTGGTAGGAGGAACACCATCTGATCTGCTTTACTAGTGTTCCCTGTGAACAAAGAAA
GTGATTCTGTGTTTCAACATCATCAATCTTTGAAA

>G1804 Amino Acid Sequence (domain in AA coordinates: 357-407)

MVTRETKLTSEREVESMAQARHNGGGGGENHPFTSLGRQSSIYSLTLDEFQHALCENGK
NFGSMNMDEFLVSIWNAEENNNNQOAAAAAGSHSV PANHNGFNNNNNNGGEGGVGVFSG
GSRGNEDANNKRGIANESSLPQGSLTLPAPLCRKTVDEVWSEIHRGGGSGNGGDSNGRS
SSSNQGNNAQNGGETAARQPTFGEMTLEDFLVKAGVVRHPTNPKPNPNQNPSSVI
PAAAQQQLYGVFQGTGDPSPFGQAMGVGDPSPGYAKRTGGGGYQQAPPVQAGVCYGGGVGF
GAGGQMGVMVGPLSPVSSDGLGHGQVDNIGGQYGVDMGGLRGRKRVVDGPVEKVVERRQR
RMIKNRESAARSARKQAYTVELEAELNQLKEENAQLKHALAELERKRKQYFESLKSRA
QPKLPKSNGLRLTLMRNPSCPL*

>G207 (16..930)

aaaagatctgttcaatggcgatcgtgttaagggtccatggagtcagaagaagatgag
cagctacgaaggatgggtgagaaatacggaccgaggaattggtctgcgattagcaaatcg
attccaggtcgatctggttaaatacgtgtagattacgttggtgtaatacagttatctccggag
gttgagcatcgtctttctcgccggaggaagatgagactattgtaaccgcccgtgctcag
tttgtaacaagtggcgacgattgctcgtcttctaaccggtcgtacggataacgccgtt
aaaaatcactggaactctacgcttaagaggaaatgcagcggaggtgtggcggttacgacg
gtgacggagacggaggaagatcaggatcgccgaagaagaggagatctgttagctttgat
cctgcttttgctccggtggatactggattgtacatgagtcctgagagtcctaaccggaatc
gatgtagtgattctagcacgattccgtcaccgtcgtctcctggtgctcagctgtttaaa
ccaatgccgatttccggcggttttacgggtggtccgcagccgttacgggttgaaatgtct
tcgtcttcggaggatccacctacttcgttgagtttgctactacctggagctgagaacacg
agttcgagccataacaataacaacaacgcgttgatgtttccgagatttgagagtcagatg
aagattaatgtagaggagagaggaggaggaggagaaggacgtagaggtgagtttatgacg
gtggtgcaggagatgataaaagctgaagtgaggagttacatggcggaatgcagaaaaca
agtgggtgattcgtcgtcggaggtttatacgaatccggcggaatggtggttttagggat
tgtggagtaataaacacctaaaggttgagtagtttgggttaggggttaaaacttgaaatcgat
tggggattttcaagagcattcatttttgggggttatggtaaaattaaaaacaaaaacaaa
atgtacagaggaattaaaatttctatggaataatcttaaatctcaaatatttgttacttg
ttttggtgattcataacccaaatcaaa

>G207 Amino Acid Sequence (domain in AA coordinates: 6-106)

MADRVKGPWSQEEDEQLRMVKEYGPRNWSAISKIPGRSGKSCRLRWCNQLSPEVEHRP
FSPPEEDETIVTARAQFGNKWATIARLLNGRTDNVKNHWNSTLKRKCSGGVAVTTVTETE
EDQDRPKRRSVSFDPAFAPVDVTGLYMSPESPNGIDVSDSSTIPSPSPVAQLFKPMPIS
GGFTVVPQPLPVEMSSSSEDPPTSLSLPGAENTSSSHNNNNNALMFPFRFESQMKINVE
ERGGGEGRRGEFMTVVQEMIKAEVRSYMAEMQKTSGGFVVGGGLYESGGNGGFRDCGVIT
PKVE*

>G218 (1..1182)

ATGGAGGCAGAGATCGTGAGACGATCGGAGGTAACGGGATTAAGAAGGGAGGTGGAAGAA
TCGTCAATTGGTAGAGGAGATTGCGATGCTGATGGCGGCGATGTGGGAGAAGATGCGGCA
GGGTTTCGTTGGGACGAGCGGGAGAGGAAGAAGAGATCGAGTTAAAGGGCCGTGGTCAAG
GAGGAGGATGATGTGTTGAGTGAGCTCGTTAAGAGGTTGGGAGCGAGGAATTGGAGTTTT
ATCGCTCGGAGTATTCCTGGTCTGTTCAAGCAAGTCTTGTCTGCTCTCGTTGGTGTAATCAG
CTCAATCCAATCTTATACGCAATTCATTACTGAGGTAGAGGATCAGGCTATCATCGCA
GCACATGCCATCCACGGAAACAAATGGGCTGTTATCGCGAAGCTCCTCCCCGGAAGACA
GATAATGCTATCAAGAACCCTGGAACCTGCTTTAAGACGTCGATTATAGACTTTGAA
AAGCCCAAGAATATAGGAACCTGGAAGCTTGGTCGTGGATGATTCTGGATTGACAGAACG
ACAACAGTAGCCTCATCAGAAGAACTTTATCTTCAGGCGGTGGTTGCCATGTAACACTACT
CCAATTGTATCTCCAGAAGGCAAAGAAGCTACCACCTCCATGGAAATGTCTGAAGAACAA
TGCGTAGAGAAAACAAACGAGAAGGTATTTCTAGGCAAGATGATAAGGATCCTCCAACG
CTTTTCGCCCCAGTGCCTCGGCTCAGTTCTTTTAAATGCTTGCAATCACATGGAAGGATCA

CCCTCTCCACATATACAAGACCAAAATCAGCTCCAATCATCTAAACAAGACGCAGCAATG
CTAAGATTGCTTGAAGGAGCTTACAGCGAACGGTTTGTGCCTCAAACATGTGGAGGTGGT
TGTTCAGCAACAATCCCGATGGCAGTTTTTCAGCAAGAATCATTGTTGGGTCCAGAGTTT
GTGGATTACTTAGACTCACCAACGTTTCCGAGTTCGAACTAGCTGCTATAGCAACGGAA
ATAGGCAGCCTCGCTTGGCTGAGAAGCGGTTTAGAGAGTAGCAGCGTGAGGGTGATGGAA
GACGCAGTTGGTCGGTTAAGGCCTCAAGGCTCCAGGGGTTCATCGAGATCATTATCTTGTA
TCTGAACAGGGGACGAACATAACCAATGTCCTGTCCACATAA

>G218 Amino Acid Sequence (domain in AA coordinates: TBD)
MEAEIVRRSEVTGLRREVEESSIGRGDCDGDGDVGEDAAGFVGTSGRGRDRVKGPWSK
EEDDVLSSELVKRLGARNWSFIARSIPGRSGKSCRLRWCNQLNPNLIRNSFTEVEDQAIIA
AHAIHGNKWAVIAKLLPGRDINAIGNHNSALRRRFIDFEKAKNIGTSLVVDSDSGFDRT
TTVASSEETLSGGGGCHVTTPIVSPEGKEATTSMEMSEEQVEKTNGEGISRQDDKDPPT
LFRPVPRLSSFNACNHMEGSPSPHIQDQNLQSSKQDAAMLRLLEGAYSERFVPQTCCGG
CCSNPDGSGFQGESLLGPEFVDYLDSPFPSSSELAIAIEIGSLAWLRSGLESSSVRVME
DAVGRRLRPQGSRGHRDHYLVSEQGTNITNVLST*

>G241 (46..867)
GAAAAAATTTCACCTCTTTTATCAGCAATCACAAATCAAAGAGATGGGAAGAGCTCCA
TGCTGTGAGAAGATGGGGTTGAAGAGAGGACCATGGACACCTGAAGAAGATCAAATCTTG
GTCTCTTTTATCCTCAACCATGGACATAGTAAGTGGCGAGCCCTCCCTAAGCAAGCTGGT
CTTTTGAGATGTGGAAGAGCTGTAGACTTAGGTGGATGAACATTTAAAGCCTGATATT
AAACGTGGCAATTTACCAAAGAAGAGGAAGATGCTATCATCAGCTTACACCAATACTT
GGCAATAGATGGTCAGCGATTGCAGCAAACTGCCCTGAAGAACCGATAACGAGATCAAG
AACGTATGGCACACTCACTTGAAGAAGAGACTCGAAGATTATCAACCAGCTAAACCTAAG
ACCAGCAACAAAAGAAGGGTACTAAACCAAAATCTGAATCCGTAATAACGAGCTCGAAC
AGTACTAGAGCGAATCGGAGCTAGCAGATTTCATCAAAACCTTCTGGAGAAAGCTTATTT
TCGACATCGCCTTCGACAAGTGAGGTTCTTCGATGACACTCATAAGCCACGACGGCTAT
AGCAACGAGATTAATATGGATAACAAACCGGGAGATATCAGTACTATCGATCAAGAATGT
GTTTCTTTTGAACCTTTTGGTGCGGATATCGATGAAAGCTTCTGGAAGAGACTGTAT
AGCCAAGATGAACACAACTACGTATCGAATGACCTAGAACTCGCTGGTTTAGTTGAGATA
CAACAAGAGTTTCAAACTTGGGCTCCGCTAATAATGAGATGATTTTGGACAGTGAGATG
GAACCTTCTGGTTCGATGATTGGCTAGAACCGGCGGGGAACAAGATCTCTTAGCCGGGCT
CTAGTTAACATGTTTGGAGAGTAAAGTGAAATGGTGCAAATTAGTTAAGGCTAAGAAATT
CAAAAGCTTTTGTTTACCGAGAAAAAACACACTCTAACTCTTGATGTGATGTAGTTAGT
GTATTAATTAGAGGCTGCGTTTTCAA

>G241 Amino Acid Sequence (domain in AA coordinates: 14-114)
MGRAPCEKMLKRPWTPPEEQILVSVFILNHGHSNWRALPKQAGLLRCGKSCRLRWMNY
LKPDIKGNFTKEEDAIISLHQILGNRWSAIAKLPGRDNEIKNVWHHLKRLLEDYQ
PAKPKTSNKKKGTKPKSESVITSSNSTRSESELADSSNPSGESLFTSPSTSEVSSMTLI
SHDGYSEINMDNKPDIIDTIDQECVSFETFGADIDESFWKETLYSQDEHNYVSNLEVA
GLVEIQQEFQNLGSANNEMIFDSEMELLVRCIG*

>G254 (15..923)
CGATTTTCGAGCTCTATGGTGTCCGTAAACCCCTAGACCTAAGGGTTTTCCAGTTTTTCGATT
CCTCGAATATGAGTTTACCAAGCTCCGATGGATTTGGTTTCGATTCCGGCCACGGGACGGA
CCAGTACGGTGTCTGTTTCTGAGGATCCGACGACGAAGATTCCGGAAGCCGTACACAATCA
AGAAGTCGAGAGAGAATTGGACAGATCAAGAGCAGGATAAATTTCTAGAAGCTCTTCACT
TATTCGATAGGGATTGGAAGAAAATAGAAGCCTTTGTTGGATCAAAAACAGTAGTTTCTAGA
TACGAAGCCACGCTCAGAAATACTTTCTCAAAGTTTCAAGAGAGTGGTGCTAACGAACATC
TTCCACTTCTCGACCTAAGAGGAAAGCGAGTCATCCTTATCCTATAAAGGCTCCTAAAA
ATGTTGCTTATACCTCTCTCCGCTCTTCGAGTACATTACCGTTGCTTGAGCCTGGTTATT
TGTATAGCTCTGATTGCAAGTCATTGATGGGAAACAGGCTGTTTGTGCATCTACCTCTT
CTTCGTGGAATCATGAATCGACAAATCTGCCAAAACCGGTGATTGAAGAGGAACCGGGAG
TCTCGGCCACGCTCCTCTCCCAAATAATCGCTGCAGACAGGAAGATACAGAGAGGGTAC
GAGCAGTGACAAAGCCAAATAACGAAGAAAGTTGTGAAAGCCACATAGAGTGATGCCGA
ATTTTGCTGAAGTTTACAGCTTCATTGGAAGTGTCTTCGATCCCAACACATCAGGCCACC
TCCAGAGATTAAAGCAGATGGAATCAATAAATATGGAACGGTCTTTTACTGATGCAAAA
ACCTGTCTGTAAATCTGACAAAGTCCCGAGTTTGCAGAGCAAAGGAGGTTGATATCATCAT
ACAGCGCTAAAGCTTTGAAATAGAGATAGAATAAAACAATAATGTACCTTATGTGAGATC

AAGAGACAATCATCCAAGGTCTGTATGCATTGCTTGGATTTAGGCCTCGTGTCTCACTA
CAGGAGCAGAACCAATCGCAAAGACTCTTAGATGGCTACTGAGTTGTGGTTTTATGTCT
CTGTAAAGTCGCGGTGGAGCACACGTGTTTGTCTGTCTTGTGTATGTGTGTATAGATAAT
ACAAGGTTTTGTCAGAGTAAGGTCACAGTTAGCTGCAAGTGAGTTGGATCAATCTTAAGA
TTAAACCCTTGAGAGTGAGTGTCCAAAGAGACTGTGTAATATTGGTTTGGCGGTCAGCAG
AAGAGTTTTGAAGTGCACATCCAGTTAGTGATAACACGGTTGAAGAAAAGGTAAGGTTAC
AAGTTTAGTTTTGAATAATTGTATACTCAAAAAATATGAATGTATAAAGAATAATCACTT
GAGTCGCCTTA

>G254 Amino Acid Sequence (domain in AA coordinates: 62-106)
MVSVNPRPKGFPVFDSSNMSLPSSDGFSGIPATGRTSTVSFSEDPTTKIRKPYTIKKSRE
NWTQDQHDKFLEALHLFDRDWKKIEAFVGSKTVVQIRSHAQKYFLKVQKSGANEHLPLPR
PKRKASHPYPIKAPKNVAYTSLPSSSTLPLEPGYLYSSDSKSLMGNQAVCASTSSSWNH
ESTNLPKPVIEEPEGVSATAPLPNNRCRQEDTERVRAVTKPNNEESCEKPHRVMPNFAEV
YSFIGSVFDPNTSGHLQRLKQMDPINMETVLLLMQNL SVNLTSPFAEQRLISSYSAKA
LK*

>G26 (73..729)
TTGGCTTGTACCCAAACCCATCTTTGACTTCAAAAATAAAAATAAAAATAATCATAATTGA
CATCATCGGATAATGCATAGCGGGAAGAGACCTCTATCACCAGAATCAATGCCGGAAT
AGAGAAGAGAAAAAGAGTTGTGTGTTGCTCAACTTTGTGCGGAATCTGATGTGTCTGAT
TTTGTCTCTGAAGTCACTGCTCAACCCATCCCATCATCCATTGATGATCAATCTTCGTCG
CTTACTCTTCAAGAAAAAAGTAAGTCAAGGCAACGAACTACAGAGGCGTGAGGCAAGA
CCGTGGGGAAAAATGGGCGGCTGAGATTCTGTGACCCGAACAAGGCAGCTCGTGTGTGGCTT
GGGACGTTTCGACACTGCAGAAGAAGCCGCTTAGCGTATGATAAAGCTGCATTTGAGTTT
AGAGGTACAAAGGCCAAGCTTAACTTCCCGAGCATATTCTGTGCAACCCCTACTCAACTC
TATCCATCGCCCGCTACTTCCCATGATCGCATTATCGTGACACCACCTAGTCCACCTCCA
CCAATTGCTCCTGACATACTTCTTGATCAATATGGCCACTTCAATCTCGAAGTAGTGAT
TCCAGTGCCAACTTGTCCATGAATATGCTGTCTTCTTCTGCTTTCATCTTGAATCATCAA
GGGCTAAGACCAAAATTTGGAGGATGGTGAAGAACGTAAGAACATTAGTATCCACAAACGA
CGAAAATAACATGTTAATGGCATAAATATCTTCTGTCCTCAAGTTATCAAACGCATTGACC
TCCGCTTTTGATCATTTTGTAGGCGCTTAATCTCTTACGACTTCATTTTGGTAGTCTTTAA
AGAGTCTATGGAGTGGATTTAGCTAGGAATCAGGCCTTATGGATGAAAAATATATAAATT
TTGAACATGACTATGCAAGATGGGATGAAGACTACTTAGCTTGAAAAACGTCCTGATAG
GTCATGACGACTATATCCACAGAAGATGACCGACGGAGACAACAACATGCCTCACCTGAT
CGACCGATCAAAATGAGATAATGTGTTGACCGGACCGGTCGGATCAGGTTGGGTCGAGTAT
ATCA

>G26 Amino Acid Sequence (domain in AA coordinates: 67-134)
MHSGKRPLSPESMAGNREEKKELCCSTLSESDVSDVSELTGQIPSSIDDQSSSLTLQ
EKSNSRQRNYRQVRQRPWGKWAABEIRDPNKAARVWLGTDFDTAEAAALAYDKAAFEFRGHK
AKLNFPEHIRVNPTQLYPSPATSHDRIIVTPSPPPPPIAPDILLDQYGHFQSRSSDSSAN
LSMNLSSSSSLNHQGLRPNLEDEGENVKNISIHKRRK*

>G263 (48..902)
TTTTTAGTTTTATTCTTCTGTGGTAAAATAAAAAAGTTTCGCCGGAGATGACGGCTGTGA
CGGCGGCGCAAAGATCAGTTCCGGCGCCGTTTTTAAGCAAACGTATCAGCTAGTTGATG
ATCATAGCACAGACGAGCTCGTTTCATGGAACGAAGAAGGAACAGCTTTTGTCTGTGGA
AAACAGCAGAGTTTGTAAAGATCTTCTTCTCAATACTTCAAGCATAATAATTCTCAA
GCTTCATTCTGTCAGCTCAACACTTACGGATTTTCGTAAACTGTACCGGATAAATGGGAAT
TTGCAAACGATTATTTCCGGAGAGGCGGGGAGGATCTGTTGACGGACATACGACGGCGTA
AATCGGTGATTGCTTCAACGGCGGGGAAATGTGTTGTTGTTGTTGCTTCTGAGTCTA
ATTCTGTTGGTGGTGTGATGATCAGGTTCAAGCTCCACGTCATACCCGGTTCGTGGAAGA
ATCCTGGTTCCGGTGGAGAATGTTGCTGATTTATCAGGAGAGAACGAGAAGCTTAAAC
GTGAAAACAATAACTTGAGCTCGGAGCTCGCGGCGGGAAGAAGCAGCGGATGAGCTAG
TGACGTTCTTGACGGGTCTCTGAAAGTAAGACCGGAACAAATCGATAAAATGATCAAAG
GAGGGAAATTTAAACCGGTGGAGTCTGACGAAGAGAGTGAGTGCGAAGGTTGCGACGGCG
GCGGAGGAGCAGAGGAGGGGGTAGGTGAAGGATTGAAATGTTTGGGGTGTGGTTGAAAG
GAGAGAGAAAAAGAGGACCGGGATGAAAAGAATTATGTGGTGTGAGTGGGTCCCGTATGA
CGGAAATAAAGAACGTGGACTTTACGCGCGGTTGTGGAAAAGCAGCAAAGTCTGCAACT
AAAAAAGAGTAGAAGACTGTTCAAACAGCGGTGTGACACGTCATCGACGACGACGAAAA

AAATGATTTAAAAAACTATTTTTTTCCGTAAGGAAGAAAAGTTATTTTTATGTTTTAAAA
AGGTGAAGAAGGTCCAGAAGGATCAACGCAAATATATAAATGGATTTTCATGTATTATAT
AATTTAATTAGTGTATTAAGAAAA

>G263 Amino Acid Sequence (domain in AA coordinates: TBD)
MTAVTAAQSRVPAFLSKTYQLVDDHSTDDVVSWEETAFVVKTAFAKDLLPQYFKH
NNFSSFIQLNTYGRKTVDPKWEFANDYFRRGGEDLLDIRRRKSVIASTAGKCVVVG
PSESNSGGDDHGSSTSSPGSSKNPGSVENMVADLSGENEKLKRENNLSSELAAAKKQ
RDELVTFLTGHLKVRPEQIDKMIKGGKFKPVESDEESECEGCDGGGGAEEGVGEGLKLF
VWLKGERKKRDRDEKNYVVSGRMTEIKNVDFHAPLWKSSKVCN*

>G308 (196..1794)
AGTAATTTAGTTTTTTTTTTTTTTTTTACAATTTATTTTGTATTAGAAAGTGGTAGTGG
AGTGA AAAACAAATCCTAAGCAGTCCTAACCGATCCCCGAAGCTAAAGATTCTTCACCT
TCCCAAAATAAGCAAAACCTAGATCCGACATTGAAGGAAAACCTTTTAGATCCATCTCT
GAAAAAAACCCAACCATGAAGAGAGATCATCATCATCATCAAGATAAGAAGACTATG
ATGATGAATGAAGAAGACGACGGTAACGGCATGGATGAGCTTCTAGCTGTTCTTGGTTAC
AAGGTTAGGTCATCGGAAATGGCTGATGTTGCTCAGAACTCGAGCAGCTTGAAGTTATG
ATGTCTAATGTTCAAGAAGACGATCTTTCTCAACTCGCTACTGAGACTGTTCACTATAAT
CCGGCGGAGCTTTACACGTGGCTTGATTCTATGCTCACCAGCCTTAATCCTCCGTCGTCT
AACGCGAGTACGATCTTAAAGCTATTCGGGTGACGCGATTCTCAATCAGTTCGCTATC
GATTCCGCTTCTTCGCTTAACCAAGGCGGCGGAGGAGATACGTATACTACAAACAAGCGG
TTGAAATGCTCAAACGGCGTCGTGGAAACACCACAGCGACGGCTGAGTCAACTCGGCAT
GTTGCTCTGGTTGACTCGCAGGAGAACGGTGTGCGTCTCGTTACCGCGCTTTTGGCTTGC
GCTGAAGCTGTTTCAGAAAGAGAACTGACTGTGGCGGAAGCTCTGGTGAAGCAAAATCGGA
TTCTTAGCTGTTTCTCAAATCGGAGCTATGAGACAAGTCGCTACTTACTTCGCGGAAGCT
CTCGCGCGGCGGATTTACCGTCTCTCTCCGTCGCAGAGTCCAATCGACCACTCTCTCTCC
GATACTCTTCAGATGCACTTCTACGAGACTTGTCTTATCTCAAGTTCGCTCACTTCACG
GCGAATCAAGCGATTCTCGAAGCTTTTCAAGGGAAGAAAAGAGTTCATGTCAATTGATTTT
TCTATGAGTCAAGGTCTTCAATGGCCGGCGCTTATGCAGGCTCTTGCCTTCGACCTGGT
GGTCTCTCTGTTTTCCGGTTAACCAGAAATGGTCCACCGGCACCGGATAATTTTCGATTAT
CTTCATGAAGTTGGGTGTAAGCTGGCTCATTTAGCTGAGGCGATTACGTTGAGTTTGGAG
TACAGAGGATTTGTGGCTAACACTTTAGCTGATCTTGATGCTTCGATGCTTGAGCTTAGA
CCAAGTGAGATTGAATCTGTTGCGGTTAACTCTGTTTTTCGAGCTTCAAGCTCTTGGGA
CGACCTGGTGCGATCGATAAGGTTCTTGGTGTGGTGAATCAGATTAAACCGGAGATTTTC
ACTGTGGTTGAGCAGGAATCGAACATAATAGTCCGATTTTCTTAGATCGGTTTACTGAG
TCGTTGCATTATTACTCGACGTTGTTGACTCGTTGGAAGGTGTACCGAGTGGTCAAGAC
AAGGTATGTCGGAGGTTTACTTGGGTAAACAGATCTGCAACGTTGTGGCTTGTGATGGA
CCTGACCGAGTTGAGCGTCACTGAAACGTTGAGTCAAGTGGAGAACCGGTTCCGGCTCTGCT
GGGTTTGGCGCTGCACATATTGGTTTCAATGCGTTTAAAGCAAGCAGTATGCTTTTGGCT
CTGTTCAACGGCGGTGAGGGTTATCGGGTGGAGGAGAGTGACGGCTGTCTCATGTTGGGT
TGGCACACACGACCGCTCATAGCCACCTCGGCTTGGAACTCTCCACCAATTAGATGGTG
GCTCAATGAATTGATCTGTTGAACCGGTTATGATGATAGATTCCGACCGAAGCCAAACT
AAATCCTACTGTTTTTCCCTTTGTCACTTGTTAAGATCTTATCTTTCATTATATTAGGTA
ATTGAAAAATTTTAATCTCGCCTAAATTACT

>G308 Amino Acid Sequence (domain in AA coordinates: 270-274)
MKRDHHHHQDKKMTMMNEEDDGNMDELLAVLGYKVRSEEMADVAQKLEQLEVMSNVQ
EDDLSQLATETVHYNPAELYTWLDSMLTDLNPPSSNAEYDLKAIPGDAILNQFAIDSASS
SNQGGGGDTYTTNKRLLKCSNGVVETTTATAESTRHVVLVDSQENGVRVHALLACAEAVQ
KENLTVAEALVKQIGFLAVSQIGAMRQVATYFAEALARRIYRLSPSQSPIDHSLSDLQOM
HFYETCPYLKFAHFTANQAILEAFQGGKRVHVIDFMSQGLQWPALMQALALRPGGPPVF
RLTGIGPPAPDNFDYLHEVGCKLAHLAEAIHVEFEYRGFVANTLADLDASMLELRPSEIE
SVAVNSVFEHLKHLGRPGAIDKVLGVVNQIKPEIFTVVEQESNHNSPIFLDRFTESLHYY
STLFDSLEGVPSGQDKVMSEVYLGKQICNVVACDGPDRVERHETLSQWRNRFSGAGFAAA
HIGSNAFKQASMLLALFNGGEGYRVEESDGCLMLGWHTRPLIATSANKLSTN*
>G38 (149..1156)
GAGGAAAACTCGAAAAGCTACACACAAGAAGAAGAAAAGATACGAGCAAGAAGACT
AAACACGAAAAGCGATTTATCAACTCGAAGGAAGAGACTTTGATTTTCAAATTCGTCCCC
TATAGATTGTGTTGTTTCTGGGAAGGAGATGGCAGTTTATGATCAGAGTGGAGATAGAAA

CAGAACACAAATTGATACATCGAGGAAAAGGAAATCTAGAAGTAGAGGTGACGGTACTAC
TGTGGCTGAGAGATTAAAGAGATGGAAAGAGTATAACGAGACCGTAGAAGAAGTTTCTAC
CAAGAAGAGGAAAGTACCTGCCAAAGGGTCGAAGAAGGGTTGTATGAAAGGTAAAGGAGG
ACCAGAGAATAGCCGATGTAGTTTCAGAGGAGTTAGGCAAAGGATTTGGGGTAAATGGGT
TGCTGAGATCAGAGAGCCTAATCGAGGTAGCAGGCTTTGGCTTGGTACTTTCCCTACTGC
TCAAGAAGCTGCTTCTGCTTATGATGAGGCTGCTAAAGCTATGTATGGTCCTTTGGCTCG
TCTTAATTTCCCTCGGTCTGATGCGTCTGAGGTACGAGTACCTCAAGTCAGTCTGAGGT
GTGTACTGTTGAGACTCCTGGTTGTGTTTCATGTGAAAACAGAGGATCCAGATTGTGAATC
TAAACCTTTCTCCGGTGGAGTGGAGCCGATGTATTGTCTGGAGAATGGTGCAGGAGAGAT
GAAGAGAGGTGTTAAAGCGGATAAGCATTGGCTGAGCGAGTTGAACATAACTATTGGAG
TGATATTCTGAAAGAGAAAGAGAAACAGAAGGAGCAAGGGATTGTAGAAACCTGTCAGCA
ACAAACAGCAGGATTTCGCTATCTGTTGCAGACTATGGTTGGCCCAATGATGTGGATCAGAG
TCACTTGGATTCTTCAGACATGTTTGATGTGATGAGCTTCTACGTGACCTAAATGGCGA
CGATGTGTTTGCAGGCTTAAATCAGGACCGGTACCCGGGAACAGTGTTCGCAACGGTTC
ATACAGGCCCGAGAGTCAACAAAGTGGTTTTGATCCGCTACAAAGCCTCAACTACGGAAT
ACCTCCGTTTCAGTTCGAGGGAAAGGATGGTAATGGATTCTTCGACGACTTGAGTTACTT
GGATCTGGAGAATAAACAACAATATGAAGCTTTTGGATTGATATTGCCTTAATC
CCACAACGACTGTTGATTCTCTATCCGAGTTTTAGTGATATAGAGAACTACAGAACACGT
TTTTTCTTGTATATAAGGTGAACGTGTATATATCGAAACAGTGATATGACAATAGAGAAGA
CAACTATAGTTTGTAGTCTGCTTCTCTTAAGTTGTTCTTTAGATATGTTTTATGTTTTG
TAACAACAGGAATGAATAATACACACTTGTGAAGCTTTAAAAAAAAAAAAAAAAAAAAA
>G38 Amino Acid Sequence (domain in AA coordinates: 76-143)
MAVYDQSGDRNRQTIDTSRKRKSRSGDGTVAERLKRWKEYNETVEEVSTKKRKVPAKG
SKKGMKMGKGGPENSRCFSRQVVRQRIWGWVAEIREPNRGSRLWLGTFPTAQEAASAYDE
AAKAMYGPLARLNFPRSDASEVTSSTSSQSEVCTVETPGCVHVKTEDPDCESKPFSGGVEP
MYCLENGAEEMKRGVKADKHWLSEFEHNYWSDILKEKEKQKEQIVETCQQQQQDLSVA
DYGWPNDVDQSHLSDSDFDVELLRLDNGDDVFAGLNQDRYPGNSVANGSYRPESQQSG
FDPLQSLNYGIPPFQLEGKDGNGFFDDLSDLLEN*
>G43 (38..643)
CTCCTGTCTGTCTAAAGAAAAAGAGAGAGGAAGAAATGGAGACTTTTGAGGAAAGCTC
TGATTGGATGTTATACAGAAACATCTATTTGAAGACTTGATGATCCCTGATGGTTTCAT
TGAAGATTTTGTCTTTGATGATCTGCTTTTGTCTCCGACTCTGGTCTCTAGAACCTTT
TAACCCGAGTTCCGAAACTGGAACCTAGTTACCTGTTCTTGATCCAGATTCTATGTCCA
AGAGATTCTGCAAATGGAAGCAGAATCATCATCATCATCAACAACAACGTCACCTGA
GGTTGAGACTGTCTCAAAACGGAAAAAAACAAAGAGGTTTGAAGAAACGAGACATTACAG
AGGCGTGAGAAGGAGGCCATGGGGGAAATTTGCAGCAGAGATTGAGATCCGGCAAAGAA
AGGATCCAGATTGGTTAGGCACTTTTGAGAGTGATATTGATGTGCAAGGGCTTACGA
CTATGCAGCTTTAAGCTCAGGGGAAGAAAAGCTGTTCTCAACTTTCCTTTGGATGCCGG
AAAGTATGATGCTCCGGTCAATTCATGCCGAAAAAGGAGGAGAACCGATGTACCACAGCC
TCAAGGAACAACAACAAGTACTTCATCATCGTCATCAAACTAATGGGGGAATAGTGATGT
TTAATTAGTATATATAGGTTAATATCTTAAGTATGTGAAGCATCATGTATAGAGCCAAGA
ACCTGTTAGACTAGTGTACTGAAAAGAACTCTTGCAAAATATGTACTAAAGAGTTCCTGT
AACAATGGAACCTCTGCGTTTTCTCTTGTCTTAAAGAGCTTAAGGTCTAGAAACAAAGT
TCTTGCTCTTCGGTTTTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
AAAAAAAAA
>G43 Amino Acid Sequence (domain in AA coordinates: 104-172)
METFEESDLVVIQKHLFEDLMIPDGFIEDFVDDTAFVSGLSLEPFNPVPKLEPSSPV
LDPDSYVQEIQLMEAESSSSSTSTTSPEVETVSNRKKTKRFEETRHYRGVRRRPWGKFAA
EIRDPKKGSRIWLGTFESDIDAARAYDYAAFKLGRKAVLNFLDAGKYDAPVNSCRKR
RRTDVPQPQGTSTSTSSSSN*
>G536 (1..768)
ATGTCGACAAGGGAAGAGAATGTTTACATGGCGAAATTAGCCGAACAAGCTGAACGTTAC
GAAGAAATGGTTGAATTCATGGAGAAAGTTGCGAAACTGTTGATGTTGAGGAACCTTCA
GTTGAAGAGAGGAATCTTCTCTGTTGCTTACAAGAACGTGATTGGAGCGAGAAGAGCT
TCGTGGAGAATCATTTCTTCGATTGAGCAGAAAGAAGAGAGCAAAGGGAACGAAGATCAT
GTTGCTATTATCAAGGATTACAGAGGAGAGATTGAATCCGAGCTTAGCAAAATCTGTGAT
GGGATTTTGAATGTTCTTGAAGCTCATCTTATTCCTTCTGCTTACCAGCTGAATCTAAA

GTGTTTTATCTTAAGATGAAGGGTGATTATCATAGGTATCTTGCTGAGTTTAAGGCTGGT
GCTGAAAGGAAAGAAGCTGCTGAAAGCACTTTGGTTGCTTACAAGTCTGCTTCCGACATT
GCCACTGCTGAGTTAGCTCCTACTACCCGATAAGGCTTGGTCTTGCACTCAACTTCTCT
GTGTTTTACTATGAAATCCTCAACTCGCCTGATCGTGCTTGCAAGCCTCGCAAAGCAGGCG
TTTGATGATGCAATCGCTGAGTTAGATACATTGGGTGAGGAATCATACAAGGACAGTACA
CTGATTATGCAGCTTCTTAGAGACAATCTCACTCTCTGGACTTCAGATATGACTGACGAA
GCAGGAGATGAGATTAAGGAGGCATCAAAGCCCGATGGTGCCGAGTAA
>G536 Amino Acid Sequence (domain in AA coordinates:226-233)
MSTREENVYMAKLAEQARYEEMVEFMKVAKTVDVEELSVBERNLLSVAYKNVIGARRA
SWRIISSIEQKEESKGNEDHVAIIKDYRGEIESELSKICDILNVLEAHLIPASPAESK
VFYLLMKMGDYHRYLAEFKAGAEKAAESTLVAYKSASDIATAELAPHTPIRLGLALNFS
VFYYEILNSPDRACSLAKQAFDDAIAELDTLGEESYKDSLIMQLLRDNLTLWTSMDTDE
AGDEIKEASKPDGAE*
>G567 (38..1273)
AAAAAGAAGAATCAGAAAGTAAAAAGAGAGCGAGCGATGAACAGTATCTTCTCCATTGA
CGATTTCTCCGATCCTTTCTGGGAACTCCTCCGATTCTCTCAATCCCGACTCTTCTAA
GCCTGTTACGGCGGATGAAGTTAGCCAGAGTCAACCGGAATGGACTTTCGAGATGTTTCT
CGAAGAGATTTCTCGTGGCGGTGAGCTCTGAGCCACTTGGTAACAACAACAACGCGAT
CGTCGGTGTTTCTTCGGCGCAATCTCTTCTTCTGTTTCCGGACAGAATGATTTTCGAGGA
TGATAGTCGATTTCTGATCGCGATTTCGGGAAATTTGGATTGTGCTGCTCCCATGACGAC
GAAGACGGTGAATGTTGATTCGATGATTATCGTCTGTTCTTAAGAACAAGCTTGAGGC
TGAGTGCGCGACTGGTGTTTCTCTTCGGGTGCGGTCTGTGAAGCCTGAAGATTGACTAG
TTCTCCAGAACTCAACTTCAACCAGTTCAATCCAGTCTCTTACTCAAGGAGAACTTGG
TGTTACTTCTTCTTACCAGCTGAGGTGAAAAAACTGGTGATCAATGAAGCAGGTAC
TAGTGATCCTCGAGAGAAATATTCTGATGACGAGGACCTTGATGAAGAGAATGAAACCAC
CGGTTCTTGAAGCCAGAGGACGTTAAAAAATCTAGAAGGATGCTGTCAAATCGTGAGTC
AGCTAGGCGATCTAGAAGGAGAAAGCAGGAGCAAAACAGTGACCTCGAAACACAGGTTAA
TGATCTAAAGGTGAGCATTCACTTCTTAAACAACCTGAGCAACATGAATCACAAGTA
TGACGAGGCTGCTGTTGGCAATAGAACTAAAGGCTGACATTGAGACATTAAGAGCTAA
GGTGAAAAATGGCGGAAGAAACCGTGAAGAGAGTAACAGGAATGAATCCGATGCTTCTCGG
AAGATCAAGTGGACATAACAACAACAGAAATGCCAATAACTGGTAACAACAGGATGGA
TTCTTCTAGCATTATTCAGCTTATCAACCACACTCAAACCTAAACCATATGTCAAACCA
AAACATCGGGATCCCAACCATTCTACCTCCAAGACTCGGAAACAATTTCTGCTGCTCCTCC
ATCCCAAACAGCTCTCCCTTGACAGAGAATTAGAAATGGGCAAAATCACCATGTTACTCC
AAGCGCCAAACCGTATGGCTGGAATACCGAACCTCAGAACGATTGAGATGGCCGAAAAA
ATGCGTGGACTGATCAACAAGAAGCGGGTTTCGCACTATATTAATGTCTATGCATCTGT
AATTTGTAAGTGTTATTAAGTTACGAATCATGAGAAAACATCTTGTGAAAATACAGTCTC
ATGGCTTATATATATATAAGCTCTGTCTTATAACATTACAAGATTCTTATTTGAGAAT
CGTCTTTCTATTTATAGCTAATAAAAAA
>G567 Amino Acid Sequence (domain in AA coordinates 210-270)
MNSIFSIDDFSDPFWETPIPLNPDSSKPVTADEVVSQSQPEWTFEMFLEISSAVSSEP
LGNNNNAIVGVSSAQSLPSVSGQND FEDDSRFRDRDSGNLDCAAPMTTKTVNVDSDDYRR
VLKNKLEAECATGVSRLRVGSVKPEDSTSSPETQLQPVQSSPLTQGELGVTSSLPAEVKKT
GVSMKQVTSGSSREYSDDEDLDEENETGSLKPEDVKSRRLSNRESARRSRRRKQEQT
SDLETQVNDLKGEHSSLLKQLSNMNHKYDEAAVGNRILKADIETLRKVKMAEETVKRVT
GMNPMLLRSSSGHNNNRMPIITGNRMDSIIIPAYQPHSNLNMNSQNIPTILPPRL
GNNFAAPPSQTSSPLQRIIRNGQNHVTPSANPYGWNTPEQNDSAWPKKCVD*
>G580 (338..2275)
CAGTTATCTTCTTCTTCTTCTTCTGTTTTTAAATTTATTTTATAGAGAATTTTTTTG
TTTTGCTTCCGATTGATTATTTCCGGGAACGATGACTTCTCCGGGAGTTCCCGGTGAG
ATGATAAGTCAGATTGCATACTGTCTCCTCCATGGCTACTCTCAAGGGTTTGGCTGCG
GTGGATTGTTTTGTTTTCTCTAGAATCTAAAGAGGTTATCACAACGGCTTTGCAATTTGA
AACTTTTCATGTTTGGGGAGATCAAAGATGGTTTTCTTTTTTATACCTTACTTGTAGAGA
GGATTGGAAGCAGCGAATAGCTGCAACCGGTCTGTTATGGATACTAATACATCTGGAGA
AGAATTATTAGCTAAGGCAAGAAAGCCATATACAATAACAAAGCAGCGAGAGCGATGGAC
TGAGGATGAGCATGAGAGGTTTCTAGAAGCCTTGAGGCTTTATGGAAGAGCTTGGCAACG
AATTGAAGAACATATTGGGACAAAGACTGCTGTTTCAGATCAGAAGTCATGCACAAAAGTT

CTTCACAAAGTTGGAGAAAGAGGCTGAAGTTAAAGGCATCCCTGTTTGCCAAGCTTTGGA
CATAGAAATTCCGCTCTCTCGTCTAAACGAAAACCCAATACTCCTTATCCTCGAAAACC
TGGGAACAACGGTACATCTTCTCTCAAGTATCATCAGCAAAAGATGCAAAACTTGTTC
ATCGGCTCTTCTTACAGTTGAATCAGGCGTTCTTGGATTGGAATAATGCCGTTCTC
TGAGAAAACATCAACTGGAAAAGAAAATCAAGATGAGAATTGCTCGGGTGTTCCTACTGT
GAACAAGTATCCCTTACCAACGAAAACAGGTAGTGGCGACATTGAAACAAGTAAGACCTC
AACTGTGGACAACGCGGTTCAAGATGTTCCCAAGAAGAACAAGACAAAGATGGTAACGA
TGGTACTACTGTGCACAGCATGCAAACTACCCTTGGCATTTCACGCAGATATTGTGAA
CGGGAATATAGCAAAATGCCCTCAAAATCATCCCTCAGGTATGGTATCTCAAGACTTCAT
GTTTCATCCTATGAGAGAAGAACTCACGGGCACGCAAACTTCAAGCTACAACAGCATC
TGCTACTACTACAGCTTCTCATCAAGCGTTTCCAGCTTGTCTTACAGGATGATTACCG
TTCGTTTCTCCAGATATCATCTACTTTCTCCAATCTTATTATGTCAACTCTCTACAGAA
TCCTGCAGCTCATGCTGCAGTACATTCTGCTGCTTCGGTCTGGCTTATGCGAGTGTGCG
GAATTCCTGGTGATTTCATCAACCCCAATGAGCTCTTCTCCTCCAAGTATAACTGCCATTGC
CGCTGCTACAGTAGCTGCTGCAACTGCTTGGTGGGCTTCTCATGGACTTCTTCTGTATG
CGCTCAGCTCCAAATACATGTGTTCCATTCTCAACTGTTGCAGTTCCAAGTCCAGCAAT
GACTGAAATGGATACCGTTGAAAATACTCAACCGTTTGAGAAAACAAACACAGCTCTGCA
AGATCAAACCTTGGCTTCGAAATCTCCAGCTTCATCATCTGATGATTGAGATGAGACTGG
AGTAACCAAGCTAAATGCCGACTCAAAAACCAATGATGATAAAATTGAGGAGGTTGTTGT
TACTGCCGCTGTGCATGACTCAAACTGCCCAGAAGAAAATCTTGTGGACCGCTCATC
GTGTGGCTCAAATACACCTTCAGGGAGTGACGCAGAACTGATGCATTAGATAAAATGGA
GAAAGATAAAGAGGATGTGAAGGAGACAGATGAGAATCAGCCAGATGTTATTGAGTTAAA
TAACCGTAAGATTAAATGAGAGACAACAACAGCAACAACAAATGCAACTACTGATTCTGTG
GAAGGAAGTCTCCGAAGAGGGTCTGATAGCGTTTCAGGCTCTCTTGTCAAGAGAAAGATT
GCCTCAAAGCTTTTCGCTCCTCAAGTGGCAGAGAATGTGAATAGAAAACAAAGTGACAC
GTCAATGCCATTGGCTCCTAATTTCAAAGCCAGGATTCTTGTGCTGCAGACCAAGAAGG
AGTAGTAATGATCGTGTGGAACATGCAAGAGTCTTAAACGAGACAGACAGGATTTAA
GCCATACAAGAGATGTTCAATGGAAGTGAAAGAGAGCCAAGTTGGGAACATAAAACATCA
AAGTGATGAAAAAGTCTGCAAAAGGCTTCGATTGGAAGGAGAAGCTTCTACATGACAGAC
TTGGAGGTAAATAAAACATCCACATTTTATCAATATCTTTAAATCTAGTGTAGTAG
TTTGCTTCTCCAATCTTATGAAAGAGACTTTTAATTTTCTTCCGAACATTCTTTTGGT
CATGTCAGGTTCTGTACCATATTACCCCATGTCTTGTCTCTTGTCTGTTGTGTATGC
TACTTGTGGTCTATATGTCATCTGCTACTACTGTTAATTAACCATTAAGCAATGGATTG
TCTTTA

>G680 Amino Acid Sequence (domain in AA coordinates: 24-70)
MDINTSGEELLAKARKPYTITKQERWTEDEHERFLEALRLYGRWQRIEHHIGTKTAVQ
IRSHAQKFFTKLEKAEVKGIPTVCQALDIEIPPPRPRKPKNTPTPRKPGNNGTSSSQVSS
AKDAKLVSASSSQLNQAFDLLEKMPFSEKSTGKENQDENCSESVSTVNKYPLPTKQVSG
DIETSKTSTVDNAVQDVPKKNKDKDNDGTTVHSMQNYPWPHFADIVNGNIAKCPQNHPS
GMVSQDFMFHPMBREETHGHANLQATTASATTTASHQAFPAHSDQDYRSFLQISSTFSNL
IMSTLLQNPAAHAAATFAASVWPYASVGNSSSTPMSSPPSITAIAAATVAAATAWWA
SHGLLPVCPAPITCVPFSTVAVPTPAMTEMDTVENTQPFQNTALQDQTLASKSPASS
SDDSDETGVTKLNADSKTNDKIEEVVVTAHVHDSNTAQKKNLVDRSSCGSNTPSGSDAE
TDALDKMEKDKEDVKETDENQPDVIELNNRKKMRDNNNNNATDSWKEVSEEGRIAFQ
ALFARERLPQSFSPPQVAENVNRKQSDTSMPLAPNFKSQDSCAADQEGVVMIGVGTCKSL
KTRQTGFKPKYKRCSMEVKESQVGNINNQSDEKVKRLRLEGEAST*

>G867 (64..1098)

CACAAACACAAACACATTTCTGTTTTCTCCATTGTTTTCAAACCATAAAAAAAACACAGAT
TAAATGGAATCGAGTAGCGTTGATGAGAGTACTACAAGTACAGGTTCCATCTGTGAAACC
CCGGCGATAACTCCGGCGAAAAAGTCTGTCGGTAGGTAACCTTATACAGGATGGGAAGCGGA
TCAAGCGTTGTGTTAGATTGAGAGAACGGCGTAGAAGCTGAATCTAGGAAGCTTCCGTCG
TCAAAATACAAAGTGTGGTGCACAAACCAACGGAAGATGGGGAGCTCAGATTTACGAG
AAACACCAGCGGTGTGGCTCGGGACATTCAACGAAGAAGACGAAGCCGCTCGTGCCTAC
GACGTCGGGTTTACAGGTTCCGTCGCGTGACGCCGTACAAAATTTCAAAGACGTGAAG
ATGGACGAAGACGAGGTGATTTCTTGAATTTCTATTGAAATCTGAGATCGTTGATATG
TTGAGGAAACATACTTATAACGAAGAGTTAGAGCAGAGTAAACGGCGTCTAATGGTAAC
GGAAACATGACTAGGACGTTGTTAACGTCGGGGTTGAGTAATGATGGTGTCTTCTACGACG

GGGTTTAGATCGGCGGAGGCACTGTTTGAGAAAGCGGTAACGCCAAGCGACGTTGGGAAG
CTAAACCGTTTGGTTATACCGAAACATCAGCGAGAGAAACATTTCCGTTACCGTCAAGT
AACGTTTCCGTGAAAGGAGTGTGTTGAACTTTGAGGACGTTAACGGGAAAGTGTGGAGG
TTCCGTTACTCGTATTGGAACAGTAGTCAGAGTTATGTTTTGACTAAAGGTTGGAGCAGG
TTCGTTAAGGAGAAGAATCTACGTGCTGGTGACGTGGTTAGTTTCAGTAGATCTAACGGT
CAGGATCAACAGTTGTACATTGGGTGGAAGTCGAGATCCGGGTCAGATTTAGATGCGGGT
CGGGTTTGTAGATTGTTCCGAGTTAACATTTACCGGAGAGTTCAAGAAACGACGTCGTA
GGAAACAAAAGAGTGAACGATACGTAGATGTTATCGTTGGTGTGTAGCAAGAAGCAACGC
ATCTTTACGCCTCGTAACAACCTCTTCTCTTTTTTTTTCTTTGTTGTTTAAATAATTT
TAAAAAATCCATTTTCGTTTTCTTTATTTGCATCGGTTTCTTCTTCTTGTTTACCAA
GGTTCATGAGTTGTTTGTGTATTGATGAACGTAAATTTTATTTATAGGATAAATTT
TAAAAAAAAAAAAAAAAAAAAA

>G867 Amino Acid Sequence (domain in AA coordinates: 59-124)
MESSSVDESTTSGSICETPAITPAKKSSVGNLYRMGSGSSVVLDSSENGVEAESRKLPS
KYKGVVPQPNRWRGAQIYEKHQRVWLGTFFNEEDEAARAYDVAVHRFRRRDAVTNFKDVKM
DEDEVDFLNSHSEIIVMLRKHTYNEELEQSKRRRNNGNMTRTLTLTSLNDGVSTTG
FRSAEALFEKAVTPSDVGKLNRLVIPKHAEKHFPLPSSNVSVKGVLNLFEDVNGKVWRF
RYSYWNSSQSYVLTGWSRFVKEKNLRAGDVVSFRSNGQDQQLYIGWKSRSGLDLAGR
VLRLFGVNISPESRRNDVVGKRVNDTEMLSLVCSKKQRIHAS*

>G956 (1..840)
ATGGAGGAGACAGAAAAGAATAAGGGCAGCATAAGTATGGTTGAGGCTAATCTACCTCCT
GGTTTTAGATTCCATCCTAGAGACGACGAGCTCGTCTGTGACTACTTAATGAGAAGAACC
GTTTCGAGCCTCTATCAACCAGTTGTCTTGATCGACGTCGATCTTAACAAATGCGAGCCT
TGGGACATTCTCAAACGGCGAGAGTGGGAGGGAAAGAATGGTACTTTTACAGCCAAAAA
GACCGTAAATACGCAACAGGCTACAGAACAACCGGGCTACGGCCACCGGTTATTGGAAA
GCCACCGGGAAGATAGAGCAATCCAAAGAAACGGTGGTCTTGTGGGTATGAGAAAGACA
CTTGTGTTTTACCGAGGTCGATCCCCTAAAGGTCGTAAAACTGATTGGGTCATGCATGAG
TTTCGCTCTCCAAGGAAAACCTTCTCACTCCCTAATTCTCTCGAGGAAGAGTGGGTA
TTGTGTAGAGTTTCCACAAGAAGCAGCAACGGAGCTGATATAGACGACATCACAAGGAGC
TGCTCTGATGCAACAGCTTCTGCATTCTGACTCTTACATCAACTTCGACCATCATCAC
ATCATCAATCAGCATGTACCCTGCTTCTCCAATAATTTGTACATAACCAAACCAACCAA
TCCGGTTTAATCTCCAAGAACTCCAGCCCATGTTTAAATGCTTCCCCTGATCAAATGATT
CTCAGAACTTTGCTAAGTCAACTCACAAGAAAGTCAAGAAATCAGAGAGTCGTGGAGAC
GGAAGCTCAGAGAGCCAATTGACCGACATTGGCATCCCAAGCCATGCATGGAATTACTGA

>G956 Amino Acid Sequence (domain in AA coordinates: TBD)
MEETEKNKGSISMVEANLPPGFRFHPRDELVDYLMRRTVRSLYQPVLIDVDLNKCEP
WDIPQATARVGGKEWYFYSQKDRKYATGYRTNRATATGYWKATGKDRAIQRNGGLVGMKRT
LVFYRGRSPKGRKTDWVMHEFRLQGLLHHSFNSLEEEWVLCRVFHKNSNGADIDITRS
CSDATASAFMDSYINFDDHHIINQHVPCFSNNLSHNQTNQSGLISKNSPLFNASPDQMI
LRTLLSQLTKKVEESQSRGDGSSESQLTDIGIPSHAWNY*

>G996 (53..1063)
CGATCGATCTTGAATTGATTCTTTGTAGTATTTTATTACATATATATATAGATGGGAAG
ACATTATGTTGTTACAAACAGAACTGAGGAAAGGACTTTGGTCTCCTGAAGAAGATGA
GAAGCTTCTTCGTTACATCACTAAGTATGGTCATGGTTGCTGGAGCTCTGTCCCTAAACA
AGCTGGTTTACAGAGATGTGAAAAAGTTGTAGATTAAAGATGGATAAATTATTTAAGACC
AGATTTGAAGAGAGGAGCATTTTCTCAAGATGAAGAAATCTCATTTATTGAAC TTCATGC
CGTCTTGGCAATAGATGGTCTCAGATAGCTGCACAGCTTCTTGGGAAGAACCGACAATGA
AATCAAGAATCTTTGGAATTTCTGTTTGAAGAAGAAATGAGGCTGAGAGGAATTGACCC
GGTTACACACAAGCTCTTAACCGAAATCGAAACCGGTACAGATGACAAAACAAACCGGT
TGAGAAGAGTCAACAGACCTACCTCGTTGAGACTGATGGCTCCTCTAGTACCACTACTTG
TAGTACTAACCAAAACAACAACACTGATCATCTTTATACCGGAAATTTTCGGTTTTCAACG
GTTAAGTCTAGAAAACGGTTCAAGAATCGCAGCCGGTCTGACCTCGGTATCTGGATTCC
CCAAACCGGAAGAAACCATCATCATATGTCGATGAAACCATCCCTAGTGCAGTGGTACT
ACCCGGTTCAATGTTCTCATCCGGTTTAAACCGGTTATAGATCTCCAATCTCGGTTAAT
TGAATTGGAAAACTCATTCTCAACCGGGCCAATGATGACAGAGCATCAGCAAATTCAGA
GAGTAAC TACAACAATTCAACATTCTTTGGAATGGGAATCTGAATTGGGGATTAAACAAT
GGAGGAAAATCAAATCCATTCAACAATATCGAATCATTCAAATTCGCTCTTATACAGTGA

TATAAAATCAGAGACCAATTTTTTTGGCACAGAGGCTACAAATGTTGGTATGTGGCCATG
TAACCAGCTTCAGCCTCAGCAACATGCATATGGCCATATATAAATCTTCTGTATATTAT
AA

>G996 Amino Acid Sequence (domain in AA coordinates: 14-114)
MGRHSCCYKQKLKGLWSPEDEKLLRYITKYGHGCWSSVPKQAGLQRCGKSCRLRWINY
LRPDLKRGAFSQDEENLIIELHAVLGNRWSQIAAQLPGRDNEIKNLWNSCLKKKLRLRG
IDPVTHKLLTEIETGTDDKTKPVEKSQQTYLVETDGSSSTTTCSTNQNNNTDHLTYGNFG
FQRLSLENGSRIAAGSDLGWIWIPQGRNHHHHVDETIPIAVVLPGSMFSSGLTGYRSSL
GLIELENSFSTGPMTEHQIQESNYNNSTFFGNGNLNWGLTMEENQNPFITISNHSNSSL
YSDIKSETNFFGTEATNVGMWPCNQLQPQQHAYGHI*

>G1946 (90..1547)
TCTCACCTATTGTAAAAATCACCAGTTTTCGTATATAAAACCTAATTTTCTCAAAATTCC
CAAAATTGACTTGGAAATCAAAAATCCGAATGGATGTGAGCAAAGTAACCACAAGCGACG
GCGGAGGAGATTCAATGGAGACTAAGCCATCTCCTCAACCTCAGCCTGCGGCGATTCTAA
GTTCAAACGCGCCTCTCCGTTTCTGAGCAAGACCTATGATATGGTTGATGATCACAATA
CAGATTTCGATTGTCTCTTGGAGTGCTAATAACAACAGTTTTATCGTTTGGAAACCACCGG
AGTTTCGCTCGCGATCTTCTTCTTAAGAACTTTAAGCATAATAATTTCTCCAGCTTCGTTA
GACAGCTTAATACCTATGGTTTTCAGGAAGGTTGACCCAGATAGATGGGAATTTGCGAATG
AAGGTTTTTTAAGAGGTGAGAAGCACTTGCTACAATCAATAACTAGGCGAAAACCTGCCC
ATGGACAGGGACAGGGACATCAGCGATCTCAGCACTCGAATGGACAGAATCATCTGTTA
GCGCATGTGTTGAAGTTGGCAAATTTGGTCTCGAAGAAGAAGTTGAAAGGCTTAAAAGAG
ATAAGAACGTCCTTATGCAAGAAGCTCGTCAGATTAAGACAGCAGCAACAGTCCACTGATA
ACCAACTTCAAACGATGGTTTCAGCGTCTCCAGGGCATGGAGAATCGGCAACAACAATTAA
TGTCATTCTTGCAAGGCAGTACAAAGCCCTCATTTTCTATCTCAATTCTTACAGCAGC
AGAATCAGCAAAACGAGAGTAATAGGCGCATCAGTGATACCAGTAAGAAGCGGAGATTCA
AGCGAGACGGCATTGTCGTAATAATGATTCTGCTACTCTGATGGACAGATAGTGAAGT
ATCAACCTCCAATGCAGGAGCAAGCCAAAGCAATGTTTAAACAGCTTATGAAGATGGAAC
CTTACAAAACCGGCGATGATGGTTTCTTCTAGGTATGGTACGTCTACTACCGAGGGAA
CAGAGATGGAGACTTCATCAAACCAAGTATCGGGTATAACTCTTAAGGAAATGCCTACAG
CTTCTGAGATACAGTCATCATACCAATTGAAACAACCTCCTGAAAATGTTTCGGCAGCAT
CAGAAGCAACCGAGAAGTGTATTCCTTACCTGATGATCTAATCTTCCCAGCTTCACTC
ATATGCTACCGGAAAATAATTAGAGAAGCCTCCAGAGAGTTTCATGGAACCAACCTGG
GAGGTTCTAGTCCATTACTAGATCCAGATCTGTTGATCGATGATTCTTGTCTTTCGACA
TTGACGACTTTCCAATGGATTCTGATATAGACCTGTGATTACGGTTTACTCGAACGCT
TACTCATGTCAAGCCCGGTTCCAGATAATATGGATTCAACACCAGTGGACAATGAAACAG
AGCAGGAACAAAATGGATGGGACAAAATAAGCATATGGATAATCTGACTCAACAGATGG
GTCTCTCTCTCTGAAACCTTAGATCTCTCAAGGCAAAATCCTTGATTTTGGGAGTTTT
TAAAGTCTTTTGGAGGTACACAGTCCCTGAGAGCAGCATATTCAT

>G1946 Amino Acid Sequence (domain in AA coordinates: 32-130)
MDVSKVTTSDGGGDSMETKPSQPQPAAILSSNAPPPFLSKTYDMVDDHNTDSIVWSAN
NNSFIVWKPPPEFARDLLPKNFKHNNFSSFVRQLNTYGFVKVDPDRWEFANEGFLRGQKHL
LQSIITRRKPAHGQGGHQRHSQHSNGQNSSVSACVEVGKFGLEEEVERLKRDKNVLMQELV
RLRQQQQSTDNQLQTMVQRLQGMENRQQQLMSFLAKAVQSPHFLSQFLQQNQQNESNRR
ISDTSKKRRFKRDGIVRNND SATPDGQIVKYQPPMHEQAKAMFKQLMKMEPYKTGDDGFL
LGNGTSTTEGTEMETSSNQVSGITLKEMPTASEIQSSSPIETTPENVSAASEATENCIPS
PDDLTLPDFTHMLPENNSEKPPESFMEPNLGGSSPLLDPLLIDDSLSFDIDDFPMDSI
DPVDYGLLERLLMSSPVPDNMDSTPVDNETEQEQNGWDKTKHMDNLTQQMGLLSPETLDL
SRQNP*

>G217 (84..2618)
cttcgttcttaccgagttccacgagcattagcttcagagaccttgaattggagtgcggtt
ggatcaaaaacagttgagcgaagatgaggattatgattaagggaggtgttggagaaca
ccgaagatgagattctcaaagccgctgatgaagtatggttaagaaccaatgggctcgga
tctcgtctctctcgttcgtaagtctgctaaacagtgtaaagctcgctggtacgagtggt
tcgatccatctatcaaaaagactgaatggaccagagaagaagatgagaagcttctacatc
ttgctaaacttctgcctactcaatggagaactattgctcctatttggtgggtcgtaaccat
ctcaatgtcttgaggtatgagaagctccttgatgcagcatgcactaaggatgaaaatt
atgatgcagcggtgatccacgaaaattacgtcctggtgagattgatccgaaccagaag

caaaagcctgctcgctcctgatccggtagacatggacgaagatgagaaagaaatgctttctg
aagcaagagctagattggctaacacgaggggaaagaaggctaaaagaaaagctagagaaa
aacaacttgaggaagctagaaggcttgccttctctgcaaaaaagaagagaactaaaagcag
ctgggattgatggaaggcataggaagaaagagaaaggaatcgactataatgcagaaa
ttccttttgaaaagagggcacctgcgggattttatgatactgcggatgaagatcgctcctg
ctgatcaagtaaaatttccaactaccattgaagaacttgaaggaaaaagaagagctgatg
tagaagcacatttacgcaacaagatggtgcaaggaataaaattgctcagagacaggatg
ctccagcagctatattgcaagcaacaagctgaatgatccggaagtgttaggaagaggt
caaagctgatgttaccaccaccgcagatttcagaccacgagctagaagaaattgctaaga
tgggctatgccagtgccttcttgcgagaatgaggagctaacagaaggcagtgctgcta
ctcgtgcacttttggcaaattactcacaacaccaaggcaaggaatgacacccatgagga
cacctcaaagaactcctgctggttaaagggtgatgctattatgatggaagcagaaaacctgg
ccagattaagagactctcagacacctttgctaggaggagaaaatcctgagttgcaccctt
ctgacttcactggggtcactccgagaaagaaggagattcaaacgcctaataccaatgttga
cccttcaatgactcctggtggtgctggtcttactccaagaattggcttgacgccatcaa
gggattgggtccttcttcttctatgacacccaaagggaactcccttcagggtgaacttcaca
ttaacgaagacatggacatgcagcaaaagtgcaaaacttgagaggcagagacgagaggaag
ctagaaggagtttactgctctggtttgactgggcttctcagccaaagaacgagtaccaa
tagttgcacaacctcctcctgaggaaagtgaagagccagaagagaaaattgaggaagaca
tgtcagacaggatagcgaggggaaaggcggaggaagaagcaagacaacaggcattgctta
agaagagatccaaggtcttgcaagagatcttctcagacccccagctgcttcattggcag
taattaggaactcgttgccttccagctgatggagacaaaagtctgttgcctcctactc
cgattgaggttgagataaaaatggtaagagaggagcttctacagttgctggagcatgata
atgcaaagtatccgcttgatgacaaagctgagaagaagaaggagccaagaacctacca
accgttctgcttctcaagttcttgaattgacgattttgatgaaaatgagctccaaggag
ctgacaaaatgataaaggaggaggggaagtcttctgtgtgtgtcaatgggacatgagaaca
agacacttgatgatttttagaagctcacaacacatgcgtgaatgatctcatgtatttcc
cactcgaagcgttacgagctctcaagttgctgggaacgcggacaaagtgcagctt
ttcaggaggagatggagaatgtgagaaaaagatggaggaggatgagaagaaggcagaac
acatgaaggccaagtacaaaacttatacaaaagggtcatgagaggaggcagagaccgtgt
ggacccaaatagaggcgacattgaagcaggctgagattgggtggaacagaagtagatgct
ttaaagcattgaagaggcaagaagagatggctgcctcttttaggaaaaagaatttgcaag
aggaagtgtataaagcaaaaggaaacagagagtaaaactgcagactcgctatgggaatatgt
tggcaatggttgaaaaagcagaggagataatggctcggtttccgagcacaggcattgaaga
aacaagaggatgttgaaattctcacaactgaaagaagctaagctagccactggagagg
aaggagacatagccatagccatggaagcttctgcataaaaacttgagttttgtattgctt
acaagttttaaggagacgtagcttgccttgcattggttaagtttttttaatatgagtcac
gactttgtaaaaagggttatgatataattctctgtttgtatgctttgcaagagtcagaat
ttgaatgcttcaggatcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
>G217 Amino Acid Sequence (conserved domain in AA coordinates: 8-67)
MRIMIKGVWKNTEDEILKAAVMKYGKNQWARISSLLVRKSAKQCKARWYEWLDPISIKKT
EWTREDEKLLHLAKLLPTQWRTIAPIVGRTPSQCLERYEKLLDAACTKDENDYDAADDP
KLRPGEIDPNPEAKPARPDPVMDDEKEMLSEARARLANTRGKKAKRKAREKQLEEARR
LASLQKRRELKAAGIDGRHRKRKRKGIDYNAEIPFEKRAPAGFYDTADEDRPADQVKFPT
TIEELEGRKRADVEAHLRKQDVARNKIAQRQDAPAILQANKLNDPEVVRKRSKLMLPPP
QISDHELEEIAKMGYASDLLAENEELTEGSAATRALLANYSQTPRQGMTMPMRTPQRTAG
KGDALMEAEENLARLRSQTPLLGGENPELHPSDFTGVTPRKKEIQTPNPMLTPSMTPGG
AGLTPRIGLTPSRDSSSFMTPKGTFPRDELHINEDMDMQSAKLERQRREARRSLRSG
LTGLPQPKNEYQIVAQPPPEESEBPEEKIEEDMSDRIAREKAEFEARQALLKKRSKVLQ
RDLPRPPAASLAVIRNSLLSADGKSSVVPPTPIEVADKMVREELLQLLEHDNAKYPLDD
KAEEKKGAKNRTNRSASQVLAIDDFDENELQEADKMIKEEGKFLCVSMGHENKTLDDFVE
AHNTCVNDLMYFPTRSAVELSSVAGNADKVAAFQEMENVRKKMEDEKKAHEMKAKYKT
YTKGHERRAETVWTQIEATLKQAEIGGTEVECFKALKRQEEMAASFRKKNLQEEVIKQKE
TESKLQTRYGNMLAMVEKAEEIMVGFRQAQALKKQEDVEDSHKLKEAKLATGEEEDIAIAM
EASA*
>G2192 (92..2971)
CGGAAAGAGATCAACCAACGATAGAGGAGAAGAAGAACTTGCATACGCAAAAAAAGTTTC

CCGGGAAAAATTCCAGAAACTGCTTTGGAAAAATGTGCGAGCCCGATGATAATTCCGCTAG
AAACGGCGTCACTACTCAACCTTCGAGGTCAAGGAGCTTCTAATGGATGTTGACGACTT
AGATCTTGACGGTTCATGGCCACTAGATCAAATCCCTTACTTATCCTCATCGAATCGCAT
GATTTCTCCGATTTTGTCTCCTCTTCTCTGAGCAGCCTTGCTCGCCTCTCTGGGCTTT
CTCCGACGGTGGAGGAAATGGTTTTCCACCACGCAACCTCCGGTGGCGATGATGAGAAGAT
CAGCTCTGTCTCCGGTGTTCCTTCTTTCCTCTCGCCGAGTATCCTCTCTTCTCCCTTA
CTCTTCTCCATCAGCAGCTGAGAACACAACAGAGAAGCATAACAGTTTCCAGTTTCCGTC
TCCATTGATGAGCCTAGTCCCACCAGAGAACACAGACAACCTACTGTGTGATCAAAGAGAG
GATGACTCAGGCGCTTCGATACTTCAAAGAATCAACCGAACACACGTTTTGGCTCAGGT
CTGGGCTCCTGTGAGAAAGAATGGTCGTGATTTGCTGACGACTTTGGGTCAACCTTTTGT
TCTTAATCCTAATGGTAATGGGCTTAATCAATACAGGATGATCTCTCTCACATATATGTT
TTCTGTGATAGTGAAAGTGACGTAGAGCTCGGACTCCCGGGTTCGAGTTTTCCGTCAGAA
ATTGCCTGAATGGACTCCAAATGTTTCAGTACTATTCCAGCAAAGAATTCTCGCGGCTTGA
TCACGCCTTGCACTACAACGTGCGTGGTACACTGGCCTTGCTGTCTTAAATCCCTCTGG
TCAGTCCGTCATAGGTGTGTGGAACCTTATAATGACCTCAGAGAAGATTCACTATGCACC
CGAAGTGGACAAAGTTTGCAAAGCCCTTGAGGCGGTAAATCTGAAAAGCTCGGAAATACT
TGATCACCACAAACACAGATATGCAATGAGAGTCGCCAAAACGCGCTTGCTGAGATTCT
CGAAGTGTGACAGTTGTATGTGAGACCCATAACTTGCCTCTCGCTCAGACTTGGGTTCC
ATGTACGATGGGAGCGTTCTTGCCAATGGTGGCGGTCTAAAGAAAACTGCACCAGCTT
TGACGGTAGCTGCATGGGTCAAATCTGCATGTCTACAACCGACATGGCCTGCTATGTCGT
GGATGCTCATGTCTGGGGCTTTAGAGATGCCTGTCTTGAACACCATCTCCAGAAAGGCCA
GGGAGTCGTGGACGAGCTTTTCTCAATGGTGGCTCATGTTTCTGCAGAGACATACCAA
GTTCTGCAAAACGCACTACCCACTAGTCCATTATGCGCTCATGTTCAAGTTGACCACTTG
TTTTGCAATATCTCTCCAGAGCTCTTACACGGGCGACGACAGTTACATTCTTGAATTTTT
TCTTCTCTCGAGTATAACAGACGACCAAGAGCAAGATTGTCTGTTGGGTTCTATTTGGT
GACAATGAAAGAACATTTTTCAGAGTCTGAGGGTTGCATCTGGGGTTGACTTTGGTGAAGA
TGACGACAAATTGTCTTTTCGAGATCATCCAAGCATTACCGGACAAGAAGGTTTCAATCAAA
AATAGAATCCATTTCGAGTTCCCTTTTCTGGTTTAAAGTCAAATGCAACAGAGACGATGTT
GATTCCTCAGCCTGTGGTTTCAGTCTTCTGATCCAGTAAATGAGAAAATCAACGTGGCCAC
TGTTAACGGTGTGGTTAAGGAGAAGAAGAAAAACAGAGAAAAAGCGTGGGAAGACTGAGAA
AACAAATCAGTCTAGATGTACTTTCAGCAGTATTTCACTGGAAGTCTCAAAGACGCTGCAAA
GAGCCTAGGAGTTTGGCCGACGACAATGAAGCGAATTTGCAGGCAACACCGGAATCTCGCG
GTGGCCATCGAGGAAGATCAAGAAAGTGAATCGTTCAATCACAAGCTGAAACGAGTCAT
CGAATCTGTTCAAGGTACTGATGGAGGCCTCGACCTGACTTCCATGGCCGTTAGTTCCAT
CCCTTGACACACGGTCAAACATCAGCACAGCCACTAACTCACCAATGGTTCCAAACC
ACCTGAGCTACCAAACACCAATAATTACCTAACCATTTGGTCAAGTGATCACAGTCCGAA
CGAGCCAAATGGTTTCGCTGAGTTACCACCAAGCAATGGTCAAGCGATCACGAACGGT
GGATGAGAGCGCTGGGACTCCAACCTCTCATGGCTCATGTGACGGTAACCAATAGATGA
ACGGAAGTCCCAAATCAAGATCCGCTCTTTCAGGGTTGGTGGATCACCCGGGCTCCTTTT
TCCACCTTATTCTAGAGATCATGATGATCTGCAGCTTCTTCGCAATGCCGAACAGGCT
TCTTGGTTCTATAGACCATTTCCGAGGAATGCTCATTGAAGACGCTGGAAGTTCAAAGA
TCTGAGAAATCTCTGCCCCACTGCAGCATTGACGATAAGTTTCAAGACACAACTGGAT
GAACAATGATAATAATAGCAACAACAACCTTATACGCTCCCCCAAAGGAAGAGGCCATTGC
AAATGTTGCATGCCGAACCATCAGGCTCAGAAATGAGAACGGTAACAATCAAAGCAAGTTA
CAAAGACGACATAATACGGTTCAGAATATCCTCGGGTTCAGGTATAATGGAATTGAAGGA
TGAAGTGGCTAAGAGGCTGAAAGTTGATGCAGGAACGTTTCGATATCAAGTATCTTGACGA
TGATAACGAATGGGTTTAAATAGCTTGTGATGCTGATCTTCAAGAATGTCTCGAGATCCC
TAGATCCTCCCGCAEGAAAATCGTAAGGCTCTTAGTTCATGATGTAACGACAAATCTAGG
GAGCTCCTGCGAGAGCACTGGAGAATTGTGACCTGATAATTCATTCGAACCTCTTTGTAA
ATAG

>G2192 Amino Acid Sequence (conserved domain in AA coordinates:600-700)

MCEPDDNSARNGVTTQPSRSRELLMDVDDLDDGWSWPLDQIPYLSNNRMISPIFVSSSS
EQPCSPLWAFSDGGNGFHHATSGDDEKISSVSGVPSFRLAEYPLFLPYSSPSAENTT
EKHNSFQFPSPLSLVPPENTDNYCVIKERMTQALRYFKESTEQHVLAQVWAPVRKNGRD
LLTTLGQPFVLNPNNGNLNQYRMISLTVMFSVDSVDELGLPGRVFRQKLPEWTPNVQY
YSSKEFSRLDHALHYNVRGTLALPVFNPSGQSCIGVVELIMTSEKIHYAPEVDKVCKALE
AVNLKSSEILDHQTQICNESRQNALAEILEVLTVCETHNLPLAQTWVPCQHGSVLANG

GGLKKNCTSFDFGSCMGQICMSTTDMACYVVDAAHVWGFDRDACLEHHLQKGQGVAGRAFLNG
GSCFCRDITKFKCTQYPLVHYALMFKLTTCTFAISLQSSYTGDDSYILEFFLPSSITDDQE
QDLLLLGSILVTMKEHFQSLRVASGVDFGEDDDKLSFEIIQALPDKKVHSHKIESIRVPFSG
FKSNATETMLIPQPVVQSSDPVNEKINVATVNGVVKEKKKTEKKRGKTEKTIISLDVLQOY
FTGSLKDAAKSLGVCPTTMKRICRQHGISRWPSRKIKKVNRSITKLKRVIESVQGTGGGL
DLTSMVAVSSIPWTHGQTSAPLNSPNGSKPPELPNTNNSPNHWSDDHSPNEPNGSPELPP
SNGHKRSRTVDESAGTPTSHGSCDGNQLDEPKVPNQDPLFTVGGSPGLLFPYSRDHVDS
AASFAMPNRLLSIDHFRGMLIEDAGSSKDLRLNLCPTAAFDDKFQDTNWMNNDNNSNNNL
YAPPKEEAIANVACEPSGSEMRTVTIKASYKDDIIRFRISGSGIMELKDEVAKRLKVDA
GTFDIKYLDLDDNEWVLIACDADLQECLEIPRSSRTKIVRLLVHDVTTNLGSSCESTGEL*
>G504 (69..1040)

CGTCGACCTCTTGACGATCATGAGACTGATTTCTGTAATAATATCGTCATTATATCAAATT
AGAAGTTGATGGAAAACATGGGGGATTCGAGCATAGGGCCGGCCATCCGCATCTCCCTC
CCGGGTTTCGGTTTACCCGACTGATGAGGAAC TAGTAGTTTATTACCTCAAGAAGAAAG
CAGATTCTGTTCACCTTCCAGTCTCAATCATCGCAGAGATTGATCTTTACAAGTTTGATC
CTTGGGAGCTTCCAAGCAAGGCGAGTTTTGGAGAGCACGAGTGGTACTTCTTAGTCCTC
GGGATCGGAAGTATCCAATGGGGTAGGCCAAACCGGCAGCAACTTCCGGTTATTGGA
AAGCAACGGGAACCGATAAACCGATATTTACGTGCAATAGTCACAAGGTTGGTGTCAAGA
AAGCGCTTGTTTTTTACGGTGGAAAGCCTCCTAAAGGGATAAAAACAGATTGGATCATGC
ATGAATATCGCCTCACTGATGGTAACCTTAGCACTGCGGCTAAGCCGCTGACTTAACCA
CGACAAGGAAAAACCACTACGGCTAGACGATTGGGTTCTATGTAGGATCTATAAGAAGA
ATAGTTCAAAAGACCAACAAATGGAGAGAGTATTACTTAGAGAGGATCTAATGGAAGGCA
TGCTCTCAAAATCATCTGCTAATTCTTCTTCTACATCAGTACTAGACAACAACGACAACA
ATAATAACAATAACGAAGAACACTTTTTTCGACGGTATGGTCGTTTCTTCAGACAAACGTT
CCTTGTGTGGTCAATACCGAATGGGCCACGAGGCCCTCAGGATCATCTTCATTCCGATCTT
TCTTATCGAGCAAGAGGTTTTCATCATACAGGTGATCTCAACAATGATACTACAATGTCT
CTTTTGTTCGATGCTTAGTGAGATTCTTCAGAGTTCGGGGTTTCATGCAAATGGTGTATA
TGGATACGACGCTCTCTCTAGCTGATCATGGGGTTTTAAGACAGGCGTTTCAGCTTCCTA
ACATGAAGTGGCACTCATAACTATATAGATATATATGTGTGTATCATATATGTATCTAT
GCAGGCCTAATATAGTTTACACATAAATCATCTGGGGCGGCCGCT

>G504 Amino Acid Sequence (domain in AA coordinates: TBD)
MENMGDSSIGPGPHLPFGFRFHPTDEELVHYLKKKADSVPLPVSIIEIDLYKFDPE
LPSKASFGHEHWYFFSPDRKYPNGVRPNRAATSGYWKATGTDKPIFTCNHSHKVGKAL
VFYGGKPPKGIKTDWIMHEYRLTDGNLSTAAKPPDLTTRKNSLRLLDDWVLCRIYKKNSS
QRPMTMERYLLREDLMEGLMLSKSSANSSSTSVLDNNDNNNNNNEEHFFDGMVVSSDKRSLC
GQYRMGHEASGSSSFGSFLSSKRFHHTGDLNNDNYNVSFVSMLEIPQSSGFHANGVMDT
TSSLADHGVLRQAFQLPNMNWHS*

>G622 (248..2620)
TCTTTCTTTCTTCAATTCGCCGTCAAAATCTTCTCTTTCTTCTTCCCCGCCGGTCTCTCA
CCAATCCTCTGATCTCTCTACACACGAACCTTTGATTTTGACCAACGTCGATGCATGTTT
ATGACTAGTCTCTTCTCAATCCTTCAATTTTCATCAATTCACGTCGATTTTCGTATCCGAT
TCGTTGTTCTAGCTCTTTGTGTGGTGTAGGGTTTTAAGATTTTGAATTTGGGGTTTGA
GTTGTGTATGTTTGAAGTCAAAATGGGGTCAAAGATGTGCATGAACGCTTCATGTGGTAC
GACTTCTACTGTGAATGAAGAAAGGTTGGCCTCTTCGATCTGGTCTTCTCGCTGATCT
CTGTTATCGTTGCGGATCTGCGTATGAGAGTTCTCTATCTGTGAACAATTTTCATAAGGA
CCAATCTGGTTGGAGGGAATGCTATTTGTGTAGCAAGAGACTACATGTGGATGCATTGC
TTCTAAGGTAACGATTGAGTTAATGGACTATGGTGGTGTGGTTGTAGTACATGTGCTTG
CTGCCATCAACTCAATTTGAACACAAGGGGTGAGAATCCAGGTGTTTTAGCAGATTGCC
AATGAAAACGTTAGCTGATAGGCAACATGTAAATGGCGAAAGCGAGGAAGAAACGAAGG
CGATCTCTTTTCTCAGCCACTAGTCATGGGCGGAGATAAAAGGGAAGAGTTTCATGCCTCA
CCGTGGGTTTTGGTAAGCTAATGAGTCCAGAAAGTACAACCACCGGCATAGGCTGGATGC
TGCTGGGGAAATGCATGAATCATCACCTTTACAGCCATCTTTAAATATGGGTTTGGCTGT
GAATCCGTTTAGCCCATCTTTTGAACCGAGGCTGTGAGGGAATGAAACACATCAGTCC
TTCTCAGTCCAACATGGTCCATTGCTCTGCTTCTAATATACTGCAAAAGCCATCAAGACC
TGCTATTTCAACTCCTCTGTGGCTAGTAAATCCGCTCAGGCGCGGATTGGAAGGCCTCC
TGTCGAAGGGCGAGGGAGAGGCCACTTGCTTCCGCGGTATTGGCCAAATATACGGATAA
AGAGGTTACAGCAGATCTCTGGAATTTGAATTTGAACATTGTACCTCTCTTTGAGAAAAC

TCTTAGTGCCAGTGATGCTGGTTCGCTAGTTCCTCCAAAAGCCTGTGCAGA
GGCATATTTTCCCTCCGATTAGTCAATCCGAAGGCATTCCCTTTGAAAATCCAAGATGTGAG
GGGTAGGGAGTGGACGTTCCAGTTCAGATATTTGGCCCAATAACAATAGTAGAATGTATGT
TTTAGAAGGTGTCCTCCATGCATACAGTCCATGATGCTACAGGCTGGTGATACAGTAAC
TTTCAGTCGGGTTGATCCTGGCGGAAAACCTAATCATGGGTTCCAGGAAGGCAGCTAATGC
TGGAGACATGCAGGGTTGTGGGCTCACCAACGGAACATCAACTGAGGACACATCATCGTC
TGGTGTAACAGAAAACCCACCCTCCATAAATGGTTCCTCGTGTATTTCACTAATACCGAA
AGAGTTGAATGGTATGCCTGAGAATTTGAACAGTGAGACTAACCGGGGCGAGGATAGGTGA
TGATCCTACACGAGTTAAAGAGAAGAAGAGAACTCGAACCATTTGGTGCAAAAAATAAGAG
ACTTCTTTTGCATAGTGAAGAATCTATGGAGCTGAGACTCACTTGGGAAGAAGCTCAGGA
CTTGCTTCGTCCTCTCTAGTGTAAAGCCTACCATCGTTGTTCATTGAGGAGCAAGAAAT
TGAAGAAATATGACGAACCTCCTGTCTTTGGAAAGAGGACTATAGTCACTACAAAACCTTC
AGGTGAACAGGAACGATGGGCAACTTGGCAGGACTGCTCTAAATGGAGAAGGTTACCTGT
AGATGCTCTTCTTTCTTTTAAATGGACATGTATAGACAATGTTTGGGATGTGAGTAGGTG
TTCATGTTCTGCACCGGAGGAGTCTGAAGGAACTTGAGAATGTTCTTAAAGTAGGTAG
AGAGCACAAGAAGAGAAGAACTGGGGAAGACAGGCAGCACAAAGTCAGCAAGAACCGTG
TGGTTTGGAGCGCACTGGCGAGTGCAGCAGTCTTAGGAGACACAATAGGCGAGCCAGAGGT
AGCGACCACGACGACAGATCCCAAGGCACAGGGCTGGATGCTCTTCGATCGTGTGCATTCA
GCCACCAAGTGGGAAAAGGTAGGCACAAGCCTACATGTGGCTGCACTGTGTGTAGCACCGT
GAAGAGAAGGTTCAAGACGCTTATGATGAGGAGGAAGAAGAAGCAGTTGGAGCGCGATGT
AACAGCAGCAGAAGATAAGAAGAAGAGGACATGGAAGTGGCTGAGTCTGATAAGAGTAA
GGAGGAGAAGGAAGTGAACACAGCGAGAATAGACCTGAACAGTGATCCATACAATAAAGA
AGATGTTGAAGCTGTTGCGGTGGAGAAAGAAGAGAGTCAAAAAAGAGCAATAGGACAGTG
TTCGGGCGTGGTGGCTCAAGACGCCAGTGTATGTTTAGGAGTTACAGAGTTAGAAGGAGA
GGGTAAGAATGTTCTGTGAAGAGCCGAGAGTTTCAAGCTGATATGGAAA

>G622 Amino Acid Sequence (domain in AA coordinates: TBD)

MFEVKMGSKMCMNASCCTTSTVEWKKGWPLRSGLLADLCYRCGSAYESSLFCEQFHKDQS
GWRECYLCSKRLHCGCIASKVTIELMDYGGVGCSTCACCHQLNLNTRGENPGVFSRLPMK
TLADRQHVNGESGGRNEGDLFSQPLVMGGDKREEFMPHRGFGKLMSPESSTTTGHRLLDAAG
EMHESSPLQPSLNMGLAVNPFSPSFATEAVEGMKHISSPSQSNMVHCSASNILQKPSRPAI
STPPVASKAQARIGRPPEGRGRGHLLPRYWPKYTDKEVQOISGNLNLNIVPLFEKTL
ASDAGRIGRLVLPKACAEAYFPPIISQSEGIPLKIQDVRGREWTFQFRYWPNNNSRMYVLE
GVTPCIQSMMLQAGDVTFTSRVDPGGKLIMSRKAANAGDMQGCGLTNGTSTEDTSSSGV
TENPPSINGSSCISLIPKELNGMPENLNSETNNGRIGDDPTRVKEKKRTRTIGAKNKRLL
LHSEESMELRLTWEEAQDILLRSPSPVKPTIVVIEEQEIEEYDEPPVFGKRTIVTTKPSGE
QERWATCDDCSKWRRLPVDALLSFKWTCIDNVWDVSRSCSAPEESLKELENVLKVGREH
KKRRTGERQAAQSQQEPCGLDALASAAVLGDTIGEPEVATTTTRHPRHRAGCSCIVCIQPP
SGKGRHKPTCGCTVCSTVKRRFKTLMRRKKKQLERDVTAAEDKKKKDMELAESDKSKEE
KEVNTARIDLNSDPYNKEDVEAFAVEKEESRKRAIGQCSGVVAQDASDVLGVTELEGEGK
NVREEPRVSS*

>G778 (50..1249)

TCTCAATAACACAAAACCTTTTAAACTAGTAAAATACACAGATTTTAGGATGAGCCAATG
TGTTCAAAACGTGCATCGATGATACTCCGGCAGCAGCCACCACCACCGTCCGCTCCAC
CACAGCCGCAGACATCCCCATATTAGACTACGAGGTAGCCGAGCTGACGTGGGAGAACGG
GCAACTAGGCTTGACGGCTTAGGTCCACCGGAGTGACGGCTTCGTCGACCAAGTACTC
CACAGCGCCGGTGGACGTTGGAGTCGATAGTGACCAAGCTACTCGCCTCCCTAACCC
TAAGCCCCACGGATGAGCTCGTCCCGTGGTTCCATCATCGCTCCTCCAGGCGCGATGGC
AATGGACGCGCTTGTCCTTGTCTCAACCTAGTACACGAGCAGCAGAGCAAGCCTGGTGG
CGTTGGCTCCACCCGGGTGGGTCATGTAGCGATGGTCGTACCATGGGCGGTGGAAAACG
AGCAAGAGTGGACCGGAGTGAGCGGGCGGGAGTCAGCGGCTGACCATGGACACTTA
CGACGTAGGTTTACCTCAACATCAATGGGCTCGCACGATAACACAATCGACGATCATGA
CTCCGTCTGCCACAGCCGCCACAGATGGAGGACGAAGAAGAGAAGAAAGCCGGAGGAAA
ATCATCAGTTTCAACCAAGAGAAGCAGAGCTGCTGTATTATCAACCAATCCGAACGTAA
GAGGAGAGATAAAATCAATCAAGAGATGAAGACTTTGCAAAAACCTGGTTCCCAATTCCAG
CAAGACGGATAAAGCATCTATGTTGGATGAAGTGATAGAGTATTTGAAGCAACTTCAAGC
ACAAGTGAGCATGATGAGCAGAATGAATATGCCTTCTATGATGCTTCTATGGCCATGCA
GCAACAACAACACTACAAATGTCTCTCATGTCCAATCCCATGGGTTTAGGGATGGGCAT

GGGGATGCCCCGTCTCGGTCTCCTCGACCTTAATTCTATGAACCGAGCTGCTGCAAGCGC
TCCTAATATCCATGCCAACATGATGCCAAACCCATTTTGGCCATGAATTGTCCATCGTG
GGATGCTTCTTCCAATGACTCTCGATTTTCAGTCTCCTCTCATCCCCGATCCTATGTCTGC
CTTTCTTGCTATGCTCTACTCAGCCAACGACGATGGAAGCGTATAGCAGGATGGCTACATT
ATATCAGCAAATGCAACAACAACTTCTCTCCTTCGATCCAAAATGATTATTACTCAA
ACACCTCTATATAGTTTACGTCTATATATGTGTAGTACATACATATATATATATTC
CATCATAATTATTTATTTATATGTATAGGCTTCTCATGAATTATGATATTATACGTATTA
CGTAAAAAA

>G778 Amino Acid Sequence (domain in AA coordinates: 220-267)
MSQCVPNCHIDDTAAATTTVRSTTAADIPILDYEVAELTWENGQLGLHGLGPPRVTAASS
TKYSTGAGGTLESIVDQATRLPNPKPTDELVPWFHHRSSRAAMAMDALVPCSNLVHEQQS
KPGGVGSTRVGS DGR T MG G G K R A R V A P E W S G G S Q R L T M D T Y D V G F T S T S M G S H D N T I
DDHDSVCHSRPQMEDEEBEKAGGKSSVSTKRSRAAAIHQSERKRRDKINQRMKTLQKLV
PNSSKTDKASMLDEVIEYLKQLQAQVSMMSRMNMPMSMLPMAMQQQQQLQMSLMSNPMGL
GMGMGMPGLGLLDLNSMNRAAASAPNIHANMMPNPFPMNCPSPWDASSNDSRFQSPFLIPD
PMSAFLACSTQPTTMEAYSRMATLYQQMQQLPPPSNPK*

>G791 (173..877)
TTTTCTTTGGGTGTTCTTCCACCAACGGCAGAAATCGATTGCGCTTAAATCTCCCCCTC
CTTTGATCTCTCTGATCGCCGCCGGGAACATTCAATTCCCGGGAGTTCAACAAAAAA
AAACTCTCCGTTTTATTTTCCCCCTTTTTCACCGGTGGAAGTTTCCGGAGATGGTGTC
ACCCGAAAACGCTAATTGGATTGTGACTTGATCGATGCTGATTACGGAAGTTTCACAAAT
CCAAGGTCCTGGTTTCTTCTGGCCTGTTTCAGCAACCTATTGGTGTTTCTTCTAACTCCAG
TGCTGGAGTTGATGGCTCGGCTGGAACCTCAGAAGCTAGCAAAGAACCTGGATCCAAAAA
GAGGGGGAGATGTGAATCATCCTCTGCCACTAGCTCGAAAGCATGTAGAGAGAAGCAGCG
ACGGGACAGGTTGAATGACAAGTTTATGGAATTGGGTGCAATTTTGGAGCCTGGAAATCC
TCCCAAAACAGACAAGGCTGCTATCTTGGTTGATGCTGTCCGCATGGTGACACAGCTACG
GGGCGAGGCCCAGAAGCTGAAGGACTCCAATTCAAGTCTTCAGGACAAAATCAAAGAGTT
AAAGACTGAGAAAAACGAGCTGCGAGATGAGAAACAGAGGCTGAAGACAGAGAAAGAAAA
GCTGGAGCAGCAGCTGAAAGCCATGAATGCTCCTCAACCAAGTTTTTCCCAGCCCCACC
TATGATGCCTACTGCTTTTGCTTCAGCGCAAGGCCAAGCTCCTGGAACAAGATGGTGCC
AATCATCAGTTACCCAGGAGTTGCCATGTGGCAGTTCATGCCCTCCTGCTTCAGTCGATAC
TTCTCAGGATCATGTCTTCTGTCCTCTGTTGCTTAATCAAGAAAAATCATCAACCGGTT
TGCTTCTTGCTTCCGCTTAAAAGAAAAGTCTCCATTTGTTTTGCTCTCCTCTCTTCTCG
GCTTCTTAGTCTTATCCTTTTGCTTTGTCGTGTTATCATCGTAACGTGTTATCTGTTGAA
CAATGATATGACATTGTAACTCCAATTGCTTCGCGCAATGTTATCTATTACATGTAA
TTTAAGTAGAGTTTGGCAAAAAAA

>G791 Amino Acid Sequence (domain in AA coordinates: 75-143)
MVSPENANWICDLIDADYGSFTIQPGFSPVQPIGVSSNSSAGVDGSAGNSEASKEPG
SKKRGRCESSSATSSKAKREKQRRDLNDKFMELGAILEPGNPPKTDKAILVDAVRMT
QLRGEAQKLKDSNSSLQDKIKELKTEKNELRDEKQRLKTEKEKLEQQLKAMNAPQPSFFP
APPMPTAFASAQGQAPGNKMVPIISYPGVAMWQFMPPASVDTSQDHVLRPPVA*

>G861 (158..880)
CTTCTTCTCCTCCTCCATCTCTTCTCTTACTCTCTCTTAAATCATCTCTCATTTCTGA
ATCTTGATCCATCAAAATCAATCCCGTTCTCGAAAGATCCATTAAAAATCAAAACCTAAGC
TCTCTCTCTTGCTTCTAGGGTTTTTTTGTTCGTTGTGATGGCGAGAGAAAAGATTCAGAT
CAGGAAGATCGACAACGCAACGGCGAGACAAGTGACGTTTTTCGAAACGAAGAAGAGGGCT
TTTCAAGAAAGCTGAAGAACTCTCCGTTCTCTGCGACGCCGATGTCGCTCTCATCATCTT
CTCTTCCACCGGAAAACCTGTTTCGAGTTCTGTAGCTCCAGCATGAAGGAAGTCTTAGAGAG
GCATAACTTGCAGTCAAAGAACTTGAGAAGCTTGATCAGCCATCTCTTGAGTTACAGCT
GGTTGAGAACAGTGATCAGCCCCGAATGAGTAAAGAAATTGCGGACAAGAGCCACCGACT
AAGGCAAATGAGAGGAGAGGAACCTCAAGGACTTGACATTGAAGAGCTTCAGCAGCTAGA
GAAGGCCCTTGAACTGGTTTTGACGCGTGTTGAAACAAAGAGTGACAAGATTATGAG
TGAGATCAGCGAACTTCAGAAAAAGGAATGCAATTGATGGATGAGAACAAGCGGTTGAG
GCAGCAAGGAACGCAACTAACGGAAGAGAACGAGCGACTTGGCATGCAATATGTAACAA
TGTGTCATGCACACGGTGGTGCTGAATCGGAGAACGCTGCTGTGTACGAGGAAGGACAGTC
GTCGGAGTCTATTACTAACGCCGGAACCTCTACCGGAGCGCTGTTGACTCCGAGAGCTC
CGACACTTCCCTTAGGCTCGGCTTACCGTATGGTGGTTAGAGATGGAACAATTCAAAGAA

GTTGATGGAGTGAGGAGAGTAATGTAAATCTTTTTAACTCGGTAGTAACAAGAGACAATG
TCTAAGTAGTGAATTCTCAAATGTTTGTGTAAGTTTCTGCCTATGGAAGAGGCTTTCATT
TTTATGATTTTCACTATGTATGATCTCTTCTCACTGCATTTCTGGTTAGTAACGGCTTGT
CACCGATAAACTTTCTCGTTATGGAAAGTTAGAATAAAAAAAAAAAAAAAAAAAAAA
>G861 Amino Acid Sequence (domain in AA coordinates: 2-57)
MAREKIQIRKIDNATARQVTFKRRRLFKKAEELSVLCDADVALIIFSSTGKLFEFCSS
SMKEVLERHNLQSKNLEKLDQPSLELQLVENS DHARMSKEIADKSHRLRQMRGEELQGLD
IEELQOLEKALETGLTRVIETKSDKIMSEISELQKKGMQLMDENKRLRQQGTQLTEENER
LGMQICNNVHAHGAESENAAVYEEGQSSSITNAGNSTGAPVDSESSDTSRLRLGLPYGG
*

>G938 (1..1755)
ATGATGATGTTTAAAGAGATGGGAATGTATGGAAACATGGATTCTTCTCTTCTCCACA
TCTCTCGATGTGTGCCATTACCACAAGCTGAACAAGAACCTGTAGTTGAAGATGTCGAC
TACACCGATGATGAGATGGATGTGGATGAGCTTGAGAAGAGGATGTGGAGAGACAAAATG
CCTTTGAAACGTCTCAAGGAGCAACAGAGTAAGTGTAAGAAGGCGTCGATGGTTCGAAA
CAGAGGCAGTCGCAAGAGCAAGCTAGGAGGAAGAAAATGTCTAGAGCCCAAGATGGGATC
TTGAAGTATATGTTGAAGATGATGGAAGTTGTAAAGCTCAAGGCTTTGTTTATGGTATT
ATTCTGAGAAGGGTAAGCCTGTGACTGGTGCTTCGGATAATTTGAGGGAATGGTGGAAA
GATAAGGTTAGGTTTATCGTAATGGTCCAGCTGCTATTGCTAAGTATCAGTCAGAGAAT
AATATTTCTGGAGGGAGTAATGATTGTAAACAGCTTGGTTGGTCCAACACCGCATACGCTT
CAGGAGCTTCAGGACACGACTCTTGGTTCGCTTTTATCGGCTTTGATGCAACATTGTGAT
CCACCGCAGAGACGGTTTCTTTGGAGAAAGGAGTTTCTCCACCTTGGTGGCCTAATGGG
AATGAAGAGTGGTGGCCTCAGCTTGGTTTACCAAATGAGCAAGGTCCTCTCTTATAAG
AAGCCTCATGATTGAAGAAAGCTTGGAAAGTCGGTGTTTAACTGCGGTGATCAAGCAT
ATGTCGCCGGATATTGCGAAGATCCGTAAGCTTGTGAGGCAATCAAAATGCTTGCAGGAT
AAGATGACGGCGAAAGAGAGTGCTACTTGGCTTGCCATTATTAACCAAGAAGAGGTTGTG
GCTCGGGAGCTTTATCCCGAGTCATGCCCTCTCTTCTTCTTCTTCTCATCATTAGGAAGC
GGTTCGCTTCTCATTAAATGATTGTAGCGAGTATGACGTTGAAGGTTTCGAGAAGGAACAA
CATGGTTTTCGATGTGGAAGAGCGGAAACCAGAGATAGTGATGATGCATCTCTAGCAAGC
TTTGGGGTTGCTAAAATGCAACATTTTCCCATAAAGGAGGAGGTCGCCACCACGGTAAAC
TTAGAGTTTACGAGAAAGAGGAAGCAGAACAAATGATATGAATGTTATGGTAATGGACAGA
TCAGCAGGTTACTTGTGGAAGATGGTCAAGTGTCTCACAGCAAAATGAATCTTGGATT
CAAGACAGGAGTTCAAGGGACAACCACAGATGGTTTGTCCATATAGAGACAATCGTTTA
GCGTATGGAGCATCCAAGTTTCATATGGGTGGAATGAACTAGTAGTTCTCAGCAACCA
GTCCAACCGATCGACCTATCGGGCGTTGGAGTTCCGGAAAACGGGCAGAAGATGATCACC
GAGCTTATGGCCATGTACGACAGAAATGTCCAAAGCAACCAACGCCTCTACTTTGATG
GAAAACCAAGCATGGTCATTGATGCAAAAGCAGCTCAGAATCAGCAGCTGAATTTCAAC
AGTGGCAATCAAATGTTTATGCAACAAGGACGAACAACGGGGTTAACAATCGGTTCCAG
ATGGTGTGTTGATTGACACCATTCAGATATGGCAGCATTGATTACAGAGATGATTGGCAA
ACCGGAGCAATGGAAGGAATGGGGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAAGATGTA
TCAATATGGTTCTGA

>G938 Amino Acid Sequence (domain in AA coordinates: 96-104)
MMMFNEMGMYGNDFFSSSTSLDVCPLPQAEQEPVVEDVDYTDDEMDVDELEKRMWRDKM
RLKRLKEQQSKCKEGVDGSKQRQSQEQARRKKMSRAQDGLKYLKMMEVCKAQGFVYGI
IPEKGPVGTGASDNLREWWDKVRFRDRNGPAAIAKYQSENNISGGSNDNCSLVGPTPHTL
LELDQDTTLGSLLSALMQHCDPPQRRFPLEKGVSPWPWPNGNEEWWPQLGLPNEQGPPPYK
KPHDLKKAWKVGVLTAVIKHMSPDIAKIRKLVRQSKCLQDKMTAKESATWLAIINQEEVV
ARELYPESCPPLSSSSSLGSGSLINDCSEYDVEGFEKEQHGFDEERKPEIVMMHPLAS
FGVAKMQHFPIKEEVATTNLEFTRKRKQNDNMVMVMDRSAGYTCENGQCPHSMNLF
QDRSRDNHQMVCPIYRDNRLAYGASKFHMGMKLVVPQQPVQPIDLSGVVPENGQKMIT
ELMAMYDRNVQSNQTPPTLMENQSMVIDAKAAQNQQLNFNSGNQMFQQGTNNGVNNRFQ
MVFDSTPFDMAAFDYRDDWQTGAMEGMGKQQQQQQQQQDVSIWF*

>G965 (73..1956)
GATTCTCTGTGATGTCTGAATCCTTACAGGATCCAAGAGCTTTGGAAAAAGATATAAT
GAATAACAAGATATGGGTTTAGCTACTACAACCTTCTTCTATGTACAAGATTATCATCAT
CACCAAGGAATCTTTCTTCTCTAATGGATTCCACCGATCATCATCAACCACTCATCAG
GAGGAAGTAGATGAATCCGCCGTCGTCTCCGGTGCTCAAATCCGGTTTATGAAACCGCC

GGAAATGTTGTCTGAAATGTTTGCTTACCCTGGCGGAGGTGGCGGCGGTTCCGGTGGAGAG
ATTCTTGATCAGTCTACTAAACAGTTGCTAGAGCAACAAAACCGTCACAACAACAACAAAT
AACTCAACTCTTCATATGTTATTACCAAATCATCATCAAGGTTTTGCTTTACCGACGAA
AACACTATGCAGCCGAGCAACAACAACACTTTACATGGCCATCTTCCTCCTCCGATCAT
CATCAAAACCGAGATATGATCGGAACCGTCCACGTGGAAGGAGGAAAGGGTTTGTCTTTA
TCTCTCTCATCTTCATTAGCCGAGCTAAAGCCGAGGAATATAGAAGCATTATTTGTGCA
GCCGTTGATGGAATCTTCTTCTTCTTAACGCATCCGCTCATCATCATCAATTCAATCAG
TTCAAGAATCTTCTTCTTGGAGAATTTCTTCTCAACATCATCACCATCAAGTTGTGGA
CATTTTGGTTCATCATCATCATCTCCCATGGCGGCTTCTTCATCCATGGAGGGATCTAC
ACGTTGAGGAATTCGAAATATACGAAACCGGCTCAAGAGTTGTTGGAAGAGTTTGTAGT
GTTGGAAGAGGACATTTCAAGAAGAACAACTTAGTAGGAACAACCTAAACCTAATACT
ACCGGTGGAGAGGAGGCGGAGGGTCCCTCGTCATCGCCGGAACAGCTAATGATAGTCCT
CCTTTGTCTCCGGCTGATCGGATTGAACATCAAAGAAGAAAAGTCAAGCTACTATCTATG
CTTGAAGAGGTGGACCGACGGTACAACCACTACTGCGAACAAATGCAAAATGGTAGTGAAC
TCATTCGACCAAGTAATGGGTTACGGCGCGGCGGTTCCGTACACGACATTAGCTCAAAAG
GCAATGTCTAGGCATTTCCGGTGTTTGAAGACGCGGTAGCGGTTAGCTTAAACGCAGC
TGTGAGCTTCTAGGGGATAAAGAGGCGGCGAGGGCTGCATCCTCGGGGTTAACCAAAGG
GAAACGCCGCGATTGCGTTTGCTAGAGCAGAGTTTGCCTCAGCAACGAGCGTTTCATCAT
ATGGGTATGATGGAGCAAGAGGCATGGAGACCGCAACGTGGTTTGCCTGAACGCTCCGTT
AATATCCTTAGAGCTTGGCTATTTCGAGCATTTCCTTAATCCGTACCCAAGCGATGCTGAT
AAGCACCTCTTAGCAGCAGAGCTGGTTTATCCAGAAATCAGGTGTCAAATTTGGTTCATA
AATGCTAGGTTTCGCTTATGGAACCAATGGTGGAGAGATGTATCAACAAGAGCAAAA
GAAAGAGAAGAAGCAGAAGAAGAAAATGAAAATCAACAACAACAAGAAGACAGCAACAA
ACAAACAACAACGACACGAAACCAACAACAATGAAAACAACCTTCACTGTCATAACCGCA
CAAACCTCCACGACGATGACATCGACACATCACGAAACGACTCTTCATTCTCTCTTCC
GTCGCCGCCGCTTCTCACGGCGGTTACGACGCGTTACCGTCGCCACGTGTCAGCAAGAC
GTCAGTGACTTCCACGTCGACGGAGATGGTGTGAACGTCATAAGATTCCGGGACCAACAG
ACTGGTGACGTGTCTCTTACGCTTGGTCTACGCCACTCTGGCAATATTCTTGATAAGAAC
ACTTCTTTCTCEGTTAGAGACTTTGGAGATTTTTAGTCTTCTTTGTTTCTCAATTTATTC
ATC

>G965 Amino Acid Sequence (domain in AA coordinates: 423-486)
MGLATTTSSMSQDYHHHQGLFSFSNGFHRSSSTTHQEVDSEAVVSGAQIPVYETAGMLS
EMFAYPGGGGGGSGGEILDQSTKQLLEQQNRHNNNNSTLHMLLPNHHQGFATDENTMQ
PQQQHQHFTWPSSSSDHHQNRDMIGTVHVEGGKGLSLSLSSSLAAAKAEYRSIYCAAVDG
TSSSSNASAHHHQFNQPKNLLLENSSSQHQQHVVGHFGSSSSSPMAASSSIGGIYTLRN
SKYTKPAQELLEEFCSVGRGHFKKNKLSRNNSNPNTTGGGGGGSSSSAGTANDSPPLSP
ADRIEHQRRKVKLLSMLLEVDRRYNHYCEQMOMVNSFDQVMGYAAVPYTTLAQKAMSR
HFRCLKDAVAVQLKRSCELLGDKEAAGAASSGLTKGETPRLRLLEQSLRQQRAPHMMGM
EQEAWRPQRLPERSVNILRAWLFEHFLNPYPADKHLARQTGLSRNQVSNWFINARV
RLWKPMVEEMYQQEAKEREEAEEENENQQQRRQQQTNNNDTKPNNNNENFTVITAQTPT
TMTSTHHENDSSFLSSVAAASHGGSDAFTVATCQQDVSDFHVDDGVNVIRFGTKQTDV
SLTLGLRHSGNIPDKNTSFSVRDFGDF*

>G1143 (54..677)

AAATAAGAAATATAAACAACCTTTTGTCTGAAAAATTATCAAAGAAGAAGAAATAAATGGGTG
GAGGAAGCAGATTTCAAGAACAGTGAGGATGAGCCGTAGGAAACAAGTAACAAAAGAGA
AGGAAGAAGATGAAAACCTCAAATCTCAAATCTGAAGCAGAGAGACGTAGAAGAGAGA
AGCTTCATTGTGCGCTTATGGCTCTGCGATCTCATGTCCCCATTGTACCAACATGACTA
AAGCAAGTATTGTTGAAGATGCGATTACTTACATAGGAGAGCTTCAAACAATGTTAAGA
ATCTCTTAGAGACATTTTATGAAATGGAAGAAGCTCCTCCTGAGATTGATGAAGAACAAA
CGGATCCAATGATAAAACCTGAAGTTGAAACTAGTGATCTTAACGAAGAGATGAAGAAC
TCGGAATCGAGGAGAATGTGCAATTGTGTAAGATTGGGGAGAGGAAGTTTGGTTAAAGA
TCATAACAGAGAAGAGAGATGGGATCTTTACTAAATTATGAGAGGTTATGAGATTTCTCG
GATTCGAGATTATCGATATTAGTCTAACAACCTCAAATGGAGCAATTCTTATTAGTGCCT
CTGTTTACAGACAGGAACCTCTGTGATGTTGAACAGACAAAAGATTTTCTTTTGAAGTTA
TGAGAAGCAATCCATAAGTATTAATTATATACATCTTGGAATTTCTTGATCTAATAACA
TTTCCATTGGTTTTTTATTACATTGTTGTTCCATTTTAAATATGATATGATTTCAGATGAAA
AAGAGTTTGTGTTACAAGCCAATGA

>G1143 Amino Acid Sequence (domain in AA coordinates:33-82)
MGGGSRFQEPVRMSRRKQVTKEKEEDENFKSPNLEAERRRRREKLHCLMALRSHVPIVTN
MTKASIVEDAITYIGELQNNVKNLLETFFHEMEEAPPEIDEEQTDPMIKPEVETSDLNEEM
KKLGIENNVQLCKIGERKFWLKIITEKRDGIFTKFMEVMRFLGFEEIIDISLTTSSNGAILI
SASVQTQELCDVEQTKDFLLEVMRNP*

>G1190 (209..2020)

TCCTGTCCCAAAACCAAAAGACTTGAGAGTGTGTCTTTAGAGAGAGATCTTCTCTCTTTT
ATCTTAGGACTCTCACTTCTTATCTCAAATCTACTTCAACTCTATTTCCAGTCTCCACAT
TTTCCCACAAATTTCAACTCTTGTCTCTTCTCCCAAAGTAAAAACAAATCGTTGCAAG
TGAGGTTTGGTTTTGGTGTATAGAATTATGAAGAGCGGGAAGCAATCTTCGCAACCTGA
AAAGGGTACTTCCAGACTCTTGTCACTGACTGTCCTGTTTTATCGCATTTTGGCGGTTTCTC
CTTCTACCTCGGTGGTATATTTTGTCTGAGAGAGACAAGATTGTAGCCAAGGATGTCAC
AAGGACGACTACAAAGGCTGTAGCTTCCCCTAAAGAACCTACAGCTACTCCTATTCAAAT
CAAATCCGTTTCTTTCCCGGAGTGGGGTCAGAGTTCCAAGATTACACCCCGTGCACCGA
TCCAAAGAGGTGGAGAAGTATGGTGTCCATCGCTTAAGTTTCTTGGAGCGTCATTGTCC
TCCGGTATATGAAAAGAAATGAGTGTGTTGATTCCACCACCAGACGGGTATAAACC GCCTAT
AAGATGGCCCAAGAGCCGAGAACAGTGTGTTGTTACAGGAACGTGCCTTATGATTGGATCAA
TAAGCAAAGTCTTAACCAGCATTGGCTTAAGAAAGAGGAGATAAGTTCCATTTCCTTGG
TGGTGGTACCATGTTCCCTCGTGGAGTTAGTCACTATGTTGATTGATGCAAGATCTGAT
TCCTGAAATGAAAGACGGAACAGTCAGGACCGCCATTGATACTGGCTGTGGGGTTGCGAG
CTGGGGAGGCGATCTTTTGGACCGTGGGATACTATCACTCTCTCTTGTCTCCAAGAGATAA
CCATGAAGCTCAGGTTCAATTTGCTCTTGAACGTGGAATTCCTGCGATTCTCGGGATCAT
CTCTACGCAACGTCTCCCTTTTCTTCAAATGCATTTGATATGGCTCATTGTTCAAGATG
TCTTATTCCTGGACAGAATTTGGTGAATCTATTTACTTGAGATTACCGTATAGTTTCG
ACCTGGAGGTTTTTTGGGTTCTTTTCTGTTCCACCTGTGAACTATAATAGACGATGGCGTGG
ATGGAACACAACCATGGAAGATCAGAAATCTGACTACAACAAGCTTCAGTCACTTCTAAC
CTCCATGTGTTTCAAAAAGTACGCTCAAAAAGATGACATAGCCGTGTGGCAGAAACTCTC
AGACAAATCTTGCTATGACAAAATCGCTAAGAACATGGAAGCTTACCTTCCCAAATGTGA
CGACAGTATAGAACCCTGATCTGTCTTGGTACACTCCACTCCGTCCTTGGCGTGGTTGCCCC
GACACCTAAAGTCAAGAAGTCTGGTCTCGGATCAATCCCAAATGGCCCCGAGAGGTTACA
TGTCGCGCCCGAGAGAATCGGTGATGTTTACGGAGGGAGTGCGAACAGTTTGAAACACGA
TGATGGTAAATGGAAGAACAGAGTTAAGCATTACAAGAAAGTTTTTACCAGCTCTTGGGAC
AGACAAGATAAGAAATGTTATGGATATGAACACTGTTTATGGAGGTTTCTCTGCGGCCCT
CATTGAGGATCCCATTGTTGGTTCATGAACGTTGTATCATCGTACAGCGCAAATTCGCTTCC
TGTTGTCTTTGATCGCGGTCTCATCGGGACTTACCACGACTGGTGCAGAGCTTTCTCAAC
GTATCCAAGACATATGATCTTCTTACCTCGACAGTCTTTTACCTTGGAGAGTCACAG
GTGTGAGATGAAGTACATTTTGCTAGAGATGGACAGGATCTTGGGCCGAGTGGATATGT
TATAATCCGAGAATCGAGTTATTTTATGGACGCAATCACACGTTAGCGAAAGGGATAAG
GTGGAGTTGCCGAGAGAGAGAGACTGAGTATGCAGTCAAAGTGAGAAGATTCTGGTTTG
CCAGAAAAAGCTATGGTTTTTCGTCAAACCAAACCTCTTGATGAGACCACCTGTATCATAG
TGTTTATCATCTCTGTGATGCACACTACAGAGAGAAGGATCTAGTCCTTTGAGTCCAAG
ATATAGCTCTATAACAATCTCCTTTTTTTTGTCTCTTTAATTTCTTGGGTATTTACCGG
TATAGATTGATATTATATATTTTTTAATTATATTTTTAATATATAGATATATAGTATGT
GGTTTAAACACTATTATTATCAAGGTCTTAAAGATTGCTTTTGAAGAGTTAAAAAATGT
TGGAGTAAGGACCTCTTGATTAATAAATTGACTGACGCAGCAAA
>G1190 Amino Acid Sequence (domain in AA coordinates: entire protein)
MKSQSSQPEKTSRILSLTVLFIACGFSFYLGIFCSEKIVAKDVTRTTTKAVAS
PKEPTATPIQIKSVSFPECGSEFQDYTPCTDPKRWKYGVHRLSFLERHCPPVYEKNECL
IPPPDGYKPIRWPKSREQCWYRNVYPYDWINKQSNQHWLKEGDKFHFPGGGTMFPRGV
SHYVDLMQDLIPEMKDGTVRTAIDTGCGVASWGGDILLDRGILSLSLAPRDNHEAQVQFAL
ERGIPAILGIIISTQRLPFPSNAFDMHCSRCLIPWTEFGGIYLLIHRIVRPGGFVWLGS
PPVYNRRWRGWNNTMEDQKSDYNKLQSLTSMCFKKAQKDDIAVWQKLSKSCYDKIA
KNMEAYPPKDDSIPEDSAWYTLRPPCVVAPTTPKVKSGLGSIKWPRLHVAPERIGDV
HGGANSLSKHDDGKWKNRVKKYKVLPA LGTDKIRNVMDMNTVYGGFSAALIEDPIWVMN
VVSSYSANSLPVVDRGLIGTYHDWCEAFSTYPRTYDLLHLSLFTLESHRCMKYILLE
MDRILRPSGYVIRESSYFMDAITTLAKGIRWSCRRETEYAVKSEKILVCQKKLWFSSN
QTS*

>G1198 (230..1675)

TCTTTTCAAATCCAATCATTGATCAACTAATCAAGAATTAATTATAAGACTTTGCAAT
CTCTCTCCCTCTCCCTCTCCCTAGCTAGTCTCTCTTGTGTTTCTTAACTCGAGCTTCTC
TCAATAGTGATTATCATCTTTTTCATCATTTCAGATTTAATGTGTTTTGCAGAAAAGAG
ACTAATCAAGAAGAGATATCATCAATTGAAGCTGTTTTCTTGAGTAGAGATGGCGAACCA
TAGAATGAGCGAAGCTACAAACCATAACCACAATCATCATCTTCCTTATTCATTATTCA
TGGTCTCAACAACAATCATCCATCTTCTGTTTTATTAAACCAAGATGGATCGTCCAGTTT
CGATTTTGGAGAGCTAGAAGAAGCAATTGTTCTGCAAGGTGTCAAGTATAGGAACGAGGA
AGCCAAGCCACCTTTATTAGGAGGAGGAGGAGGAGCTACGACTCTGGAGATGTTCCCTTC
GTGGCCAATCAGAATCACCAACTCTTCTTACTGAGAGTTCCAAGTCAGGAGGAGAGAG
CAGCGATTCAAGGATCGGCTAATTTCTCCGGCAAAGCTGAAAGTCAACAACCGGAGTCTCC
TATGAGTAGCAACATCATCTCATGCTTCAACCTCATCATAATAACATGGCAAACCTCAAG
TTCAACATCTGGAGCTTCTTCCACTTCTCGAACTTTAGCTCTCTTAAACCTTCGGAAGA
TAAGAGGAAGGCTACAACCTTCAGGCAAACAGCTTGATGCTAAGACGTTGAGACGTTTGGC
CCAAAATAGAGAAGCTGCTCGCAAAGCCGTCTTAGGAAAAAGGCGTATGTGCAACAGCT
AGAATCAAGTAGGATAAAGCTTTCCCAATTGGAGCAAGAACTTCAGCGAGCTCGTTCTCA
GGGGCTGTTTCATGGGTGTTTGTGGACCACCAGGACCTAACATCACTTCCGGAGCTGCAAT
ATTTGACATGGAATATGGGAGATGGCTAGAGGATGATAACCGGCATATGTCCGAGATTCTG
AACCAGTCTTCAGGCTCATTTATCTGACAATGATTTAAGGTTGATCGTTGACGGTTACAT
TGCTCATTTTGTAGAGATATTCGGATTAAAGCCGTGGCAGCGAAAGCCGATGTTTTTCA
CCTCATCTATTGGGACATGGATGTCCCGAGCCGAACGTTGTTTTATTGGATGGCTGGTTT
CCGTCCATCCGACCTAATCAAGATATTGGTGTGCGAAATGGATCTATTGACGGAGCAACA
ACTGATGGGAATATATAGCCTACAACACTCGTCGCAACAAGCAGAGGAGGCTCTCTCGCA
AGGCCTCGAAACAACCTTCAGCAATCTCTCATCGATACTCTCGCCGATCTCCAGTCATTGA
CGGAATGCAACAATGGCTGTGCGCTCTCGGAAAGATCTCTAATCTCGAAGGCTTTATCCG
CCAGGCTGATAACTTGAGGCAGCAGACCGTTCCAGCTGAGGCGGATCTTGACCGTCCG
ACAAGCTGCACGGTGTTCCTAGTCATCGGAGAGTACTATGGACGGCTCAGAGCTCTTAG
CTCCCTTTGGTTGTACGCCCCACGAGAGACACTGATGAGTGATGAAACCTCTGTCAAAC
GACGACGGATTTGCAGATTGTTTCAGTCATCTCGGAACCACTTCTCCAATTTCTGAATGGA
ATGAAACTTTGTATAACTAAAGGCCAAGTTTCATTGTCTGTCTGTAATTTACCTATTTC
CTTTAAAGTTGTACTAGAGAAAAGATAGGATCTTCTCTTCG

>G1198 Amino Acid Sequence (domain in AA coordinates: 173-223)

MANHRMSEATNHNHNLHPYSLIHGLNHNHPSSEGFINDGSSSFDFGELEEAIVLQGVKY
RNEEAKPPLLGGGGGATTLPEFSPWPIRTHQTLPTSSKSGGESSDSGSANFSGKAESQQ
PESPMSSKHHLMLQPHNNMANSSSTSGLPSTSRTLAPPKPSEDKRKATTSGKQLDAKTL
RRLAQNRREARKSLRKKAYVQQLSSRIKLSQLEQELQRARSQGLFMGGCGPPGPNITS
GAAIFDMEYGRWLEDDNRHMSEIRTGLQAHLSNDNLRRLIVDGYIAHFDEIFRLKAVAACA
DVFHLLIGTWMSPAERCFIWMAGFRPSDLIKILVSQMDLLTEQQLMGIYSLQHSSQQAEE
ALSQGLEQLQSLIDLTAASPVIDGMQMAVALGKISNLEGFIRQADNLRQOTVHQLRRI
LTVRQAARCFVLIGBYGRRLRALLSSLWLRSRPRETLMSDETSCQTTTDLQIVQSSRNHFSN
F*

>G1226 (212..1159)

CTGCATTTATTAAAGACAGTTTAGAAAGTGTCACCCCTAAAGGAATGTTTTAGTTTAG
AGGAAAGAGAGAGAAGAAGAGCAGCAGCAGAAGTTGTTAATTTGAAGACTATTTGAGGA
AAGACACCTATATCTAAATACTCAAAGTTACAAAATATTACTTCAGAAAACAGTTCCAT
TAGAGAGACTCATAAAGCTTCTCATCTAATTATGAGTGGATTGATGAGTTTTGGTGAATT
AGAAGACCAATTTGGTCAGATTTTCAGACACTACTATGGAAGAGAAGATACCATTTCTGCA
AATGCTTCAATGCATAGAACACCCTTTTACAACAACAGAACCAATCAGTTTCTCCAATC
ACTTCTCCAGATCCAAACCCTAGAATCAAAGAGCTGTCTCACCTTGAAACAAACATCAA
AAGAGATCCGGGTCAAACAGATGACCCGAAAAGGATCCAAGAACAGAAAACGGAGCAGT
AACGGTCAAAGAAAAAAGAAAACGGAAACGTACAAGAGCTCCAAGAACAAGACGAAGT
TGAAAACCAAAGGATGACTCACATTGCCGTGCAACGTAATCGAAGACGACAAATGAACGA
ACACTTAACTCTCTCCGATCTCTCATGCCCTCCTTCGTTTCTTCAACGGGGTGACCAAGC
TTCGATTGTAGGAGGGGCAATAGATTTTCATCAAGGAAGTACAGCAACTCTTGCAATCTCT
AGAAGCTGAGAAAACGAAGGATGGAACGTATGAAACTCTTAAACGGCGTCTGTGTTCTTC
ATCTTCGTCTCTTGATGCACTAACTCTTCTATTTCTAGCGTGTCTACGACGTCGGAAAA
TGGATTTACGGCGAGATTCGGCGGTGGAGATACGACAGAAGTGGAGGCTACGGTGATACA

GAACCATGTGAGCTTAAAAGTTTCGGTGTAAGAGAGGAAAAACGACAGATCTTAAAAGCTAT
TGTCCTCGATTGAAGAACTAAAGCTTTCGCGATTCTACATCTCACTATCTCTTCTTCTTGA
CTTTGTCTACTCTCTTCAATCTCAAGATGGAAGATGGTTGTAAATTAGGATCAGCAGA
TGAGATAGCGACAGCCGTTTCATCAGATCTTCGAGCAAATCAACGGTGAAGTCATGTGGTC
AAATCTTAGTCGAACTTAGTTGACTTTTGACTCCTAGTAACGTGTGTAACTTTAGGTTA
CAAAGAAAAGGACGTGATATAAATAAGAAAAACCAAGAGGTGAAATTTTGGGAGTTT
AATTATTATCTTATACTTTTGGATTTTAGATTAGTAGCAAACCTCGCAGTGTCTACGAT
GACATTATTATTGGTCACATGAAGGTTTAGGTTAAAAAAA
>G1226 Amino Acid Sequence (domain in AA coordinates:115-174)
MSGLMSPGELEDQFGQISDITMEEKIPFLQMLQCIEHPFTTTEPNQFLQSLQLQITLESK
SCLTLETNIKRDGPQTDDEKDPRTENGAVTVKEKRKRKRTRAPKNKDEVENQRMTHIAV
ERNRRRQMNELNSLRSLMPPSFLQRGDQASIVGGAIDFIKELEQLLQSLAEKRDGTD
ETPKTASCSSSSLACTNSSISSVSTTSNGFTARFGGDTTEVEATVIQNHVSLKVRCK
RGKRQILKAIVSIEELKLAILHLTISSSFDFVIYSFNLKMEDGCKLGSADIELATAVHQIF
EQINGEVMWSNLSRT*
>G1451 (124..2559)
TTTGTACTTCCGGAGCTAAAGAGTTATAGCTACTGTAGTAGCTGGAAGTGAAGAAGATTT
TTTAATAGATTGTACGGAATAATTAGGGTTTCAAAGTTTGGTTTCTTGAAGTTGAATTA
GACATGAAGCTGTCAACATCTGGATTGGGTCAACAGGGTCATGAAGGAGAGAAGTGTCTG
AATTCTGAGCTATGCGATGCTTGTGCTGGACCATTAGTCTCTCTCCATCATCTGGTAGT
CGAGTTGTTTACTTTCCACAGGGTCACAGTGAACAGGTAGCTGCTACAATAAGGAA
GTTGATGGTCACATACCCAATTACCAAGCCTACCACCACAATTGATATGCCAGCTCCAT
AATGTTACAATGCATGCAGATGTTGAGACGGATGAAGTCTATGCTCAAATGACACTTCAA
CCATTGACACCGGAGGAGCAGAAGGAAACATTTGTACCGATTGAGTTGGGGATACCGAGT
AAGCAACCTAGTAATTATTTTGTAAAGACTCTCACAGCTAGTGATACCAAGTACACATGGA
GGGTTTTCTGTTCCCTAGACGTGCTGCTGAGAAAGTGTTCCTCCATTGGATTACACACTG
CAGCCACCAGCTCAAGAAGTATTGCAAGGGATCTCCATGATGTTGAATGGAAGTTTAGG
CATATCTTTCCGGGACAGCCCAACGGCATCTCCTAACTACTGGATGGAGTGTCTTTGTC
AGTGCCAAAGCGACTAGTAGCTGGAGATTCTGTCTATTTTCATCAGGAATGAAAAGATCAA
CTCTTTTGGGAATTCGTCATGCCACTCGGCCGCGAGACTATTGTACCATCATCTGTTT
TCTAGTGATAGCATGCATATTGGACTCCTTGCTGCTGCTGCACATGCTTCTGCAACTAAT
AGCTGTTTCACTGTTTCTTTTCATCCAAGGGCTAGCCAATCTGAGTTTGTGATACAACCT
TCCAAGTACATTAAAGCCGTTTTTTCACACGCGTATTTTCAGTTGGGATGCGCTTTCGCATG
CTCTTCGAGACAGAAGAGTCGAGTGTCCGAGGTACATGGGTACTATAACTGGTATTAGT
GATCTAGATTCTGTTTGGTGGCCAACTCTCATTGGCGATCTGTGAAGTTGGTTGGGAT
GAATCGACTGCAGGGGAGAGACAGCCAAAGGGTTTCTTTATGGGAGATTGAGCCTCTGACT
ACCTTTCTCTATGTATCCATCTCTTTTCTCTCAGACTAAAACGTCATGGCATGCTGGC
ACATCATCTTTGCTGATGGAAGGGGTGATTGGGAAGTGGTCTAACATGGCTAAGAGGG
GGAGGTGGAGAGCAGCAAGGTTTGGCTTCTCTAAATTATCCATCTGTGGTTTGTTCCT
TGGATGCAACAAAGGCTGATCTCAGTCAAATGGGGACTGATAATAATCAGCAATACCAA
GCAATGTTAGCTGCTGGGTGTCAGAACATCGGCGGTGGAGATCCTTTAAGACAGCAGTTT
GTACAGCTGCAAGAGCCTCACCACCAATATCTTCAACAATCAGCTTCCCATATTCTGAT
TTGATGCTTCAGCAGCAACAGCAGCAACAGCGTCACGCCATCTCATGCATGCTCAAACA
CAGATTATGAGTGAGAATCTTCCGCAGCAGAATATGCGACAAGAAGTTAGTAACCAACCA
GCTGGACAGCAGCAACAGCTACAGCAACCGGACCAAAATGCATATCTTAATGCTTTCAA
ATGCAAAATGGCCATCTTCAACAGTGGCAGCAGCAATCAGAGATGCCATCTCCCTCGTTC
ATGAAGTCAGATTTTACTGACTCAAGCAACAAATTTGCAACAAC'TGCTAGTCCGGCTTCT
GGAGATGGCAATCTTTTGAATTTTCTATAACCGGTGAGTCTGTACTCCCTGAGCAGTTA
ACAACAGAGGGGTGGTCTCCAAAAGCATCCAACACTTTTCTGAACCGTGTCACTTCCA
CAAGCCTATCTTGGGAAGAGTCTTGCTCTAGAACCCGGAATCCGCAGAAATCCCTCTCTT
TTCGGTGTGATCCCGACTCTGGACTCTTCTCTCCAGTACGGTTCCCGCTTTGCTTCT
TCATCAGGAGATGCTGAAGCTTCCCCTATGTCACTAACAGATTGAGATTGAGAAATCC
TTATATAGCTGCATGCAAGACACAACATCATGAGTTATTGATGAGCTGGACAGATTAACT
TCGTCCAACCAACCAAGAACTTTGTAAAGTTTATAAACTGTTGTTTGGTGGGCGTTCA
TTAGACATCTCCCGATTGAGCAGCTACACGAGCTGCGAGAAGAGTTAGGGAAGATGTTT
GCTATCGAAGGGTGTGTTGGAAGACCCCTTAGATCAGGCTGGCAGCTTGTATTGCTTGAC
AAGGAAAATGATATCTTCTCTTGGTGATGACCCATGGGAGTCATTTGTGAATAACGTT

TGGTACATAAAGATACTATCACCAGAAGATGTGCATCAAATGGGAGATCATGGAGAAGGC
AGTGGTGGGTATTCCCGCAAACCCGACCCATCTCTAGAAGCTGCTTCGGTGTAGTCT
CATCATGCTACAACGCGGGAGCCCTTTGTTTCCCATTGGAAGTCGTTTCCACTCATCTTT
ATATGCCATTTCGTTCCGATCTCTCTCGTTTTGACGTTTTTAGAAAGAAACATAATCATAT
TTGTGAGTATGGGTCCTGAAACTTTAGGACGTACTTTAGCTTGTATTAGACAGACACTCT
CGTCATAAACATAAGAACCCTTTATGTAGCTGTCTCAGGGTAACTAACTTTTCTAG
>G1451 Amino Acid Sequence (domain in AA coordinates: 22-357)
MKLSTSGLGQOGHEGEKCLNSELWHACAGPLVSLPSSGSRVVYFPQGHSEQVAATTNKEV
DGHIPNYPSPPLQLICQLHNVTMHADVETDEVYAQMTLQPLTPPEQKETFFVPIELGIPSK
QPSNYFCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYTLQPPAQELIARDLHDVIEWKFRH
IFRGQPKRHLLTGWVSFVSAKRLVAGDSVIFIRNEKNQLFLGIRHATRPQTIVPSSVLS
SDSMHIGLLAAAAHASATNSCFTVFFHPRASQSEFVQLSKYIKAVFHTRISVGMRFRL
FTEESSVRRYMGTTIGISDLSVRWPNNSHWSVKVWGDESTAGERQPRVSLWEIEPLTT
FPMYPSLPLRLKRPWHAGTSSLPDGRGDLGSLTWLRGGGGEQQGLPLNYPVSVGLFPW
MQQRDLDSQMGTNNQYQAMLAAGLQNIIGGGDPLRQOFVQLQEPHQQYLQSSASHNSDL
MLQQQQQQQASRHLMAQTQIMSENLPQONMRQEVSNQPAQQQQQLQPPDQAYLNAFKM
QNGHLQWQQQSEMPSPSFMKSDFTDSSNKFATTASPGDGNLLNFSITGQSVLPEQLT
TEGWSPKASNTFSEPLSLPQAYPGKSLALEPGNPQNPSLFGVDPDSGLFLPSTVPRFASS
SGDAEASPMSLTDSGFQNSLYSCMQDTTHELLHGAGQINSSNQTKNFVKVYKSGSVGRSL
DISRFSSYHELREELGKMFAIEGLLEDPLRSGWQLVFVDKENDILLGDDPWESFVNNVW
YIKILSPEDVHQMDHGEHSGGLFPQNPTH*
>G1478 (1..354)
ATGTGTAGAGGGTTTGAGAAAGAAGAAGAGAGAAGAAGCGACAATGGAGGATGCCAAAGA
CTATGCACGGAGAGTCAAAAGCTCCGGTAAGCTGTGAGCTTTGCGGCGAGAACGCCACC
GTGTATTGTGAGGCAGACGCAGCTTTCTTTGTAGGAAATGCGATCGATGGGTCCATTCT
GCTAATTTTCTAGCTCGGAGACATCTCCGGCGCGTGATCTGCACGACCTGTGGAAGCTA
ACTCGTCGATGCTTGTCTGGTGATAATTTTAAATGTTGTTTTACCGGAGATAAGGATGATA
GCAAGGATTGAAGAACATAGTAGTGATCACAAAATTCCTTTGTGTTTCTCTGA
>G1478 Amino Acid Sequence (domain in aa coordinates: 32-76)
MCRGFEKEEERRSDNGGQRLCTESHKAPVSCELCGENATVYCEADAFLCRKCDRWVHS
ANFLARRHLRRLVICTTCKRLTRCLVGDNFNVVLPEIRMIARIEBHSSDHKIPFVFL*
>G1496 (116..1123)
AAACCCACCAAATAACTCAGAGCTTTTTTGCATTTTTTCCCATTCTCTATTTTGTGTTTGT
ACTTTTGGTCTCATTAAAGATCATAAGTTGAAAGATTCTGCAGAGAACAAATATGTT
GGAAGGTCTTGTCTCTCAAGAAAGCTTGTCTTAAACTCTATGGACATGTCTGTACTTGA
AAGGCTTAAATGGGTACAACAGCAACAACAGCAACTGCAACAAGTTGTGTCCCATAGCAG
TAATAATTCACCTGAACCTTCTCAGATACTTCAGTTCCATGGAAGCAACAATGATGAGTT
GTTGGAGAGTAGTTTTCAGCCAATTTCAAATGCTTGGATCTGGTTTTTGGACCAAACATAA
CATGGGTTTTTGGTCTCCACATGAATCCATTTCAAGAACAAAGTAGCTGCCATATGGAACC
TGTGGATACAATGGAGTTTTTGTGAAGACCGGTGAAGAAACAGAGCCGTTGCCTTGAA
GAACAAGAGAAAACAGAGGTTAAGACAAGGGAAGAGCAAAAGACAGAGAAGAAGATCAA
AGTAGAGGCTGAGACAGAGTCAAGCATGAAAGGAAAATCAAACATGGGAAACACTGAAGC
ATCTTCAGACACTTCAAAGGAGACATCGAAAGGAGCTTCAGAGAATCAGAAATTAGATTA
TATCCACGTGAGAGCTCGTCGAGGCCAAGCCACTGACAGACACAGCTTAGCAGAAAGGGC
GAGAAGAGAAAAGATCAGCAAGAAAATGAAATATCTGCAAGATATTGTGCCTGGATGCAA
TAAGGTCACAGGAAAAGCTGGTATGCTTGATGAGATCATCAATTATGTTCAATGTCTCCA
AAGACAAGTCGAGTTCTGTGATGAACTTGCTGTCTTGAACCCGGAAGCTAGAGCTTGC
CGTGGAAAGATGTATCGTAAACAGGCTTACTTTACAAATGTAGTTGCTTCAAAGCAATC
AATAATGGTTGATGTGCCATTGTTTCCGTTAGACCAGCAAGGATCTCTAGATTTGTCTGC
GATAAACCCGAACCAAACGACATCTATCGAAGCTCCATCTGGAAGCTGGGAAACTCAATC
ACAGAGTCTCTACAACACATCTAGCCTCGGTTTTTCATTACTAAGCAAGATTCTATTGCTCAT
AACATGGTTGACATCAATCAATCATCAAATCAGAAGCAAATCTATTACATTTGCTCAT
CAAAGTAGTAATTTGCAAATTTGGTTAATGCATTATCCTTTGATCCTTTGTTTTCTGATAT
TTAAACCAGAAGAACTGGAGATAGCAATCCAATGATCTTGTCACCA
>G1496 Amino Acid Sequence (domain in AA coordinates: 184-248)
MLEGLVSQESLSLNSMDMSVLERLKWVQQQQQLQVSHSSNNSPELLQILQFHGSNND
ELLESSFQFQMLGSGFGPNYMGFGPPHESISRTSSCHMEPVDMEVLLKTGEETRAVA

LKNKRKPEVKTREEQKTEKKIKVEAETESSMKGKSNMGNTEASSDTSKETSKGASENQKL
DYIHVRARRGQATDRHSLAERARREKISKMKYLQDIVPGCNKVTGKAGMLDEIINYVQC
LQRQVEFLSMKLAVLNPELELAVEDVSVKQAYFTNVVASKQSIMVDVPLFPLDQQGSLLDL
SAINPNQTTTSIEAPSGSWETQSQSLYNTSSLGFHY*

>G1526 (1..3090)

ATGGGAACGAAAGTCTCAGACGATCTTGTTCACCGTCAGATCAGTCGTGGGTTCGGAT
TACTCAGATATGGATATAATCAGGGCTTTACACATGGCGAATCATGATCCAACGGCTGCT
ATCAATATAATCTTCGACACTCCAAGTTTCGCCAAACCTGATGTAGCCACTCCTACCCCG
AGCGGCTCTAATGGAGGGAAGCGAGTTGATAGTGGATTAAAGGGCTGTACTTTTGGTGAC
AGCGGAAGTGTTCGAGCGAATCATCGCGTGGAGGAAGAAAATGAGAGTGTAAATGGTGGA
GGAGAAGAGAGTGTTCAGGGAATGAGTGGTGGTTTGTGGTTGTTCGAATTGGCTGGG
TTATCGACATGTAAAGGAAGGAAATTGAAGTCTGGTGATGAATTGGTGTTCACGTTTCCG
CATAGTAAAGGATTAAAGCCTGAGACTACGCCCTGGGAAGCGCGGTTTGGGCGGGGAAGG
CCAGCTTTGCGTGGTGCTTCTGATATCGTTAGGTTCTCTACAAAGGATTTCAGGAGAGATT
GGTAGAATACCAAACGAGTGGGCTCGGTGTCTTCTACCACTTGTGAGAGACAAGAAAATT
AGGATAGAAGGCAGTTGCAAGTCCGCCCTGAAGCTTTGAGCATCATGGATACAATTCTT
CTGTCTGTAAGCGTGTACATTAATAGTTCATGTTTCAAAGCATAGTGGACATTCATTT
AAGACAGCTAGTAATACGGCAGAGGAATCAATGTTCCATCCTCTCCCAATCTCTTCGG
TTACTCGGTTTGTATCCCTTTAAGAAGGCAGAGTTTACTCCAGAGGATTTTACTCTAAG
AAGCGACCTTTGAGTTCCAAGGATGGTTCTGCTATTCTACTTCGTTGCTTCAATTAAAC
AAGGTCAAGAATATGAATCAAGATGCAAACGGAGATGAAATGAGCAGTGTATCAGCGAT
GGTGATCTTGATAACATTGTTGGTGTGGGGACAGTTCTGGATTAAAGGAAATGGAACT
CCACATACACTTCTGTGTGAGCTTCGTCCATACCAAAGCAGGCACCTTCATTGGATGACC
CAACTGGAGAAAGGAAATTGCACTGTATGAGGCAGCAACAATGCTTCACCCGTGTTGGGA
GCATCTGTTTAGCAGACAAGAGGGAAGTGGTTGTCTACCTGAATCTTTTACTGTGTAT
GCTACAATACACTTCCCTAGCACACTTCAAATGGCAAGAGGAGGAATATTAGCAGACGCA
ATGGGTCTTGGAAGAGACTGTAATGACCATATCCCTTTTGCTTGCCCATTTTGGAAGCT
GCATCAACTGGGTTTCTATGCCCAACTATGAAGGAGACAAAGTGATCAGCAGTTCTGTA
GATGATCTCACTAGTCCCCCGGTGAAGGCAACCAATTCTTAGGCTTGATAAGAGGCTT
CTTGAACAAAAAAGTGTACTTCAAATGGTGGTAACCTGATTGTATGTCCGATGACACTT
TTAGGACAGTGGAAGACAGAGATTGAAATGCATGCAAAGCCTGGGTCTCTATCTGTCTAT
GTTCACTATGGGCAAAGCAGGCCGAAGGATGCAAACTTCTTTCCAGAGTGTATGGTA
ATCACCACATATGGAGTTCTAACATCCGAATTCTCGCAAGAGAACTCAGCAGACCATGAA
GGAATTTATGCAGTTTCGATGGTTTAGGATTGTTCTTGACGAGGCACATACCATCAAAAC
TCAAAAAGCCAAATTTCTTGGCTGCTGCAGCTCTGGTTGCTGATAGGCGTTGGTGTCTT
ACGGGTACTCCTATTTCAGAACAACTCTGGAGGATTTATACAGCCTTCTACGGTTTTTGAGG
ATTGAACCATGGGGAACCTTGGGCATGGTGAATAAACTTGTCCAAAAGCCATTTGAAGAG
GGTGATGAGAGAGGGTTAAAGCTAGTGCAGTCTATCTTAAACCTATCATGCTTAGGAGA
ACAAAGTCTAGCACAGACCAGAAGGAAGGCCGATTCTTGTCTACCCCTGCTGATGCA
CGGGTCATTTACTGTGAACCTTCGGAGTCTGAGAGGGATTCTACGACGCGCTATTTAAA
AGATCCAAGGTCAAATTTGATCAATTTGTTGAACAAGGCAAAGTCTTTCATAACTATGCT
TCGATCCTGGAACTGCTTTTTCGCTCTTCGACAATGTTGTGATCACCCATTTTATGTAATG
AGTCGAGGGGATACAGCGGAATACTCTGATCTGAATAAGCTTCTTAAACGTTTCCTTAGT
GGAAAGTCTTCTGGCTTAGAAAGGGAAGGAAAAGATGTACCGTCAGAGGCTTTTGTTCAG
GAGGTGGTAGAGGAAGTGCAGCAAAGGAGAGCAAGGAGAGTGTCCAATATGCCTTGAAGCA
CTTGAGGATGCTGTATTACGCCATGTGCTCATAGATTATGTCGTGAGTGTCTCTTGGCA
AGTTGGAGAAATTCTACTTCTGGGTTATGTCCTGTGTGTAGGAACACTGTAAGCAAACAA
GAACTCATCACAGCAACCAACCGAAAGTAGATTCCAGGTTGACGTGGAAGAATGGGTG
GAATCATCGAAATCACTGCTCTTCTGGAAGAGCTTGAAGGTCTTCGTTCTTCAGGCTCT
AAGAGCATTCTCTTTAGCCAGTGGACCGCTTTCTCTCGATCTCTCCAAATCCCTCTCT
CGGAATAACTTTTCAATTTGTCCCTCTTGATGGCAGCTAAGTCAGCAGCAACGAGAGAAG
GTCCTTAAAGAATTTCCGAAGATGGCAGTATCCTGGTACTGTTGATGTCTCTAAAAGCT
GGTGGCGTTGGGATAAATCTAACAGCTGCGTCCAATGCTTTTGTATGGATCCATGGTGG
AACCAGCGGTAGAGGAACAAGCTGTTATGCGTATTCATCGTATAGGGCAAACCTAAGGAA
GTCAAAATCAGAAGATTCATCGTTAAGGGAACGTTGAAGAGAGAATGGAGGCGGTTTCAG
GCGAGGAAGCAGAGAATGATCTCTGGGGCTTTAACCGATCAAGAAGTACGAAGTGCACGT
ATAGAGGAACCTCAAGATGTTATTTACCTGA

>G1526 Amino Acid Sequence (domain in AA coordinates: 493-620, 864-1006)

MGTKVSDDLVSTVRSVVGSDYSDMDIIRALHMANHDPATAINIIFDTPSFAKPDVATPTP
SGSNGGKRVDSGLKGCTFGDSSVGANHRVEEENESVNGGGEESVSGNEWFWFVGCSELAG
LSTCKGRKLKSGDELVFTTFPHSKGLKPETTPGKRGFGRGRPALRGASDIVRFSTKDSGEI
GRIPNEWARCLLPVLRDKKIRIEGSCSAPEALSIMDTILLSVSVYINSSMFQKHSATSF
KTASNTAEESMFHPLPNLFRLLGLIPFKKAEFTPEDFYSKKRPLSSKDGSAIPTSLQLN
KVKNMNQDANGDENEQCISDGLDNIIVGVDSSGLKEMETPHTLLCELRPYQKQALHWM
QLEKGNCTDEAATMLHPCWEAYCLADKRELVVYLNSTGDATIHFPSTLQMARGGILADA
MGLGKTVMTISLLLAHSWKAASTGFLCPNYEGDKVISSVDDLTSPPVKATKFLGFDKRL
LEQKSVLQNGGNLIVCPMTLLGQWKTEIEMHAKPGSLSVYVHYGQSRPKDAKLLSQSDVV
ITTYGVLTSEFSQENSADHEGIYAVRWFRIVLDEAHTIKNSKSQISLAAAALVADRRWCL
TGTPIONNLEDLYSLRLFLRIEPTWGTWAWWNKLVQKPFEEGDERGLKLVQSILKPIMLRR
TKSSTDREGRPIILVLPADARVIYCELSERDFYDALFKRSKVKFDQFVEQGVVHNYA
SILELLRLRQCCDHPFLVMSRGTAEYSDLNKLSKRFSLGKSSGLEREGKDVPEAFVQ
EVVEELRKGEQGECPICLEALEDAVLTPCAHRLCRECLLASWRNSTSGLCPVCRNTVSKQ
ELITAPTESRFQVDVEKNWVESKITALLEELEGRLSSGSKSILFSQWTAFLDLLQIPLS
RNNFSFVRLDGLTSQQQREKVLKEFSEDGSLVLLMSLKAGGVGINLTAASNAFVMDPWW
NPAVEEQAVMRIHRIGQTKEVKIRRFIVKGTVEERMEAVQARKQRMISGALTDQEVRSAR
IEBLKMLFT*

>G1543 (1..828)

ATGATAAACTACTATTTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAATATAAAGGCATCGTCACGCTTCAAGTT
TGTCCTTTTTATATTAACTGAGAGTTTTCTCTCAAACCTTACCTTTCTTCTCGATC
CTAGCTCTTAAAGAACCTTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGGAAC
TCTTCAAACCTGGATCTTACTATCTCCGTTCCAGGCTTCTCTCATCCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAGTCACGATGATGGCTCTGCTCCTCCGCGAAAGAAA
CTCCGCTTAACAGAGAACAGTCACGCTCTTCTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAACAAAAGGAAGTACTTGCCAAGCATTGTATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAACCCGTAGAGCAAGGAGCAAAATGAAGCAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTTCATTAACGGAAGAAAACACAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTGCGCTCGAGCCTT
ACTATGTGTCTCGCTGCGAGCGAGTTACCCCTGCCGCGAGCCCTTCGAGGGCGGTGGTG
CCGGTTCGGCTAAGAAAACGTTTCCGCCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)

MIKLLFTYICTYTYKLYALYHMDYACVMKYKGIVTLQVCLFYIKLRVFLSNFTFSSSI
LALKPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDEGSGGGRDQLRLDMNRLPSS
EDGDDEEFSHDDGSAPPRKKRLRLTREQSRLLDSFRQNHNLNPKQKEVLAKHLMLRPRQI
EWWFQNRARRSKLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKKTFFPQERDR*

>G162 (101..619)

AGACATACAACACCAAAATCTTCTTCTTCAACACATATTCACCTTTCACAGCAAAAAA
ACGAGAGGTTCTCTCTTATTCGTACCGTTTATAGCAAACAAATGGGTCGGAGAAAGATCAA
GATGGAGATGGTTTCAGGACATGAACACACGACAGGTTACCTTTTCAAACCGAGGACTGG
TTTGTTCAGAAGGCGAGCGAGTTAGCCACGCTCTGCAACGCTGAGTTGGGCATCGTTGT
CTTTTCAACAGGAGGCAAGCCTTTCTCTACGGGAAACCGAATCTTGATTCTGTTGCAGA
GCGATTCAAGAGAAATATGATGATTAGACAGTGGCGATGAAGAAAAAAGTGTAATTA
CAGGCCTAAACTGAAGAGGCTGAGTGAACGCTCTCGATTGCTCAACCAAGAGGTTGAAGC
TGAGAAGGAACGAGGCGAGAAGAGTCAGGAGAAGCTTGAATCTGCTGGGGATGAGAGATT
CAAGGAGTCCATTGAGACGCTTACCTCGATGAACCTCAATGAATACAAAGATAGGCTTCA
GACAGTCCATGGTAGGATTGAAGGTCAAGTCAATCACTTGCAGGCTTCGCTTGCCTCAT
GCTTCTCTCCAGAAAATAGCTAGACCGACTTGTAGAGTTACATTCTATTTTTTGTATCA
GCCTACAGAACTTACCAACACATGAAAGTTATTGCTGGTGTAGAATTTTCTGTCTATAT
GGGGTGTGACTTTCTATTGACATCAAATGAAATGTACCTGGAAATTTGTCTGTATTAA
TCTCAAGTGTAATTGCTAAACTTGATCAGCTTTTTCGCAAAAAA

>G162 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRIKIMEMVQDMNTRQVTFSKRRITGLFKKASELATLCNAELGIVVFSPGGKPFPSYK

NLDSVAERFMREYDDSDSGDEEKSGNYRPKLKRLSERLDLLNQVEAEKERGEKSQEKLE
SAGDERFKESIETLTLDLNEYKDRLQTVHGRIEGQVNHQASSCLMLLSRK*

>G1640 (168..1196)

TTCCGCCAGATCCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGTTTCGCTGACA
AGCTGCTCTAGCTTATCTGGTACCGTCGACCTCTCACTCAAGGGTCCAAAAGTGTTCCT
CTTTTTTCAGTTTCTCTTTCTCTTTTTGACAGAAGAGACCGAGAAGCAATGGGAAGGGCTC
CGTGTGTGAGAAAATCGGGTTGAAGAGAGGGAGATGGACAGCCGAGGAAGATGAGATCC
TCACCAAGTATATTTCAGACCAATGGTGAAGGTTCTTGGCGATCTTTGCCTAAGAAAGCTG
GATTGTTGAGATGTGGAAGAGCTGTAGACTAAGGTGGATAAACTACTTAAGAAGAGACT
TAAAAAGAGGAAATATTACTTCCGACGAAGAAGAAATAATCGTCAAGTTGCATTCCCTTC
TCGGCAACAGATGGTCACTTATTGCAACACATCTACCAGGAAGAACAGACAACGAAATTA
AAAATATTGGAACTCACATCTCAGCCGCAAAATCTATGCCTTCACTGCCGTTTCCGGAG
ATGGACACAATCTACTCGTCAACGATGTAGTCTTGAAGAAATCTTGTTTCATCGTCTTCTG
GAGCCAAGAACAATAACAAGACCAAGAAGAAGAAGGAAGGACTAGTAGGTCATCCA
TGAAGAAACACAAGCAAAATGGTGACGGCTCACAATGTTTCTCACAACCTAAGGAGCTAG
AGAGTGATTTTCAGTGAGGGAGGGCAAAATGGTAATTTTGAAGGAGAGTCTTTGGGGCCTT
ATGAGTGGTTGGATGGTGAGTTAGAACGGCTCTTGAGTAGTTGTGTCTGGGAATGCACTA
GTGAAGAGGCTGTGATTGGAGTAAATGATGAAAAGGTGTGTGAGAGTGGGGACAATAGTA
GTTGTTGTGTTTAATTTGTTTGAAGAAGAAACAAGGAAGCGAGACAAAGATTGGTCACGTAG
GAATCACAGAGGTTGATCATGATATGACGGTGGAAGAGAAAGAGAGGGGAAGTTTTTTAA
GTTCAATTCAAATGAAAATAATGATAAAGATTGGTGGGTTGGTCTATGTAATTCCTCAG
AAGTTGGGTTTGGGGTTGATGAGGAGTTGCTTGATTGGGAGTTTCAAGGTAATGTCACTT
GTCAAAGTGATGATCTATGGGATCTCTCAGATATTGGAGAGATAACATTGGAGTGATTGT
ACCGAGCAAGTGGATTGGCGGCCGCTCTAGACAGGCCTCGTACCGGATCTCTAGCTAGAG
CTTTCGTTTCGTATCATCGGTTTCGACAACGTTTCGTCAAGT

>G1640 Amino Acid Sequence (domain in AA coordinates: 14-115)

MGRAPCCEKIGLKRGRWTAEEDEILTKYIQTNGEGSWRSLPKKAGLLRCGKSCRLRWINY
LRRDLKRGNIISDEEEIIVKLHSLGNRWSLIATHLPGRTDNEIKNYWNSHLRKYAFT
AVSGDGHNLLVNDVVLKKSCSSSSGAKNNNKTKKKKKGRTRSMMKHKQMVTASQCFSQ
PKELESDFSEGGQNGNFEGESLGPYEWLDGELERLLSSCVWECTSEEAVIGVNDKVCES
GDNSSCCVNLFEFEEQSGSEIKIGHVGITEVDHDMTVEREREGRSFLSSNSNENNDKDWVGL
CNSSEVGFVDEELLDFEQNVTCQSDDLWDLSDIGEITL*

>G1644 (1..348)

ATGAAATTGATTGATTGGAAAGACTGTGCTTTGATGACTTACACCGAACTCATTTTGGGT
TTCTGCAATGTTTTAATGTTGATCTGCAGGAGGACTAGTGGACCTATGAGACGAGCAAAA
GGTGGTTGGACTCCAGAGGAGGATGAGACACTTAGACGAGCAGTTGAAAAGTATAAGGGG
AAGAGGTGGAAGAAAATAGCGGAATTTTCCAGAGAGAACACAAGTCCAATGCTTGCAC
AGGTGGCAGAAAGTTCTTAATCCAGAGCTTGTTAAAGGACCTTGGACTCAAGAGGTTCTC
TTATCATTTTCATGTTCTGAAACTTTTTTTGGTTTTTCATTTTACGTAA

>G1644 Amino Acid Sequence (conserved domain in AA coordinates:39-102)

MKLIDWKDCALMTYTELILGFCNVLMILICRRTSGPMRRAKGGWTPEEDELRRAVEKYKG
KRWKKIAEFFPERTQVQCLHRWQKVLNPELVKGPWTQEVLLSFSCSETFFGFHFT*

>G1646 (34..786)

GATCTTTTGATCCAATCACAAGGCAAAGATCCAATGGACAATAACAACAACAACAAC
CAGCAACCACCACCAACCTCCGCTCTATCCACCTGGCTCCGCCGTCACAACCGTAATCCCT
CTCCACCATCTGGATCTGCATCAATAGTCACCGGAGGAGGAGCGACATACCACCACCTC
CTCCAGCAACAACAGCAACAGCTTCAAATGTTCTGGACATACCAGAGACAAGAGATCGAA
CAGGTAAACGATTTCAAAAACCATCAGCTCCCTCTAGCTCGTATCAAAAAAATCATGAAA
GCTGATGAAGATGTGCGTATGATCTCCGCCGAAGCACCGATTCTTTCGCGAAAGCTTGT
GAGCTTTTCATTCTCGAATTACGATTAGATCTTGGCTTCACGCTGAAGAGAACAAACGT
CGTACGCTTCAGAAAAACGATATCGCTGCTGCGATTACTAGAACCGATATCTTCGATTTT
CTTGTTGATATTGTTCTAGGGAAGAGATCAAGGAAGAGGAAGATGCAGCATCGGCTCTT
GGTGGAGGAGGTATGGTTGCTCCCGCCGAGCGGTGTTCTTATTATTATCCACCGATG
GGACAACCGCGGTTCTCGGAGGGATGATGATTGGAAGACCGGCGATGGATCCTAGCGGT
GTTTATGCTCAGCTCCTTCTCAGGCATGGCAAAGCGTTTGGCAGAATTCAGCTGGTGGT
GGTGATGATGTCTTATGGAAGTGGAGGAAGTAGCGGCCATGGTAATCTCGATAGCCAA
GGGTAAGTGAATTCAGTAG

>G1646 Amino Acid Sequence (domain in AA coordinates: 72-162)
MDNNNNNNNQPPPTSVYPPGSAVTTVIPPPSGSASIVTGGGATYHLLQQQQQLQMF
WTYQROEIEQVNDFFKNHQLPLARIKKIMKADEDVRMISAEAPILFAKACELFILELTIRS
WLHAEENKRRTLQKNIDIAAAITRTDIFDFLVDIVPREEIKEEBDAASALGGGGMVAPAAS
GVPIYYPPMGQPAVPGGMMIGRPAMDPSGVYAQPPSQAWQSVWQNSAGGGDDVSYGSGGS
SGHGNLDSQG*

>G1672 (239..1399)

CCATTCTGACGTCCGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCTCTA
TATAAGGAAGTTCATTTCATTTGGAGAGGACACGCTGACAAGCTGACTCTAGCAGATCTG
GTACCGATCACTCCCGTCTTTATCAAATTCTTCTTCTCTTACATTTTCCCTATCCAATC
GATCTCACGCAGATCTGATCAATTCTCATCAAATCATTTAGAGATCAAAAGAAAATAT
GAAGAATAGTAAATGTAACCTCATAGATTCAAAGCTCGAAGACATCATCATCTTTGCGG
ATCAAAACATTGTCTGGATGTGGTTCGCATGATTCAAGCTGCTACTAAACCAAATTGGGT
TGGATTGCCCGGACAGGAGTGAAATTCGATCCGACAGATCAAGAACTTATAGAACATTTAGA
AGCAAAAGTGAAGGGAAAAGAAGAAAATAAGAAATGGTTCGTCTCATCCACTTATAGA
TGAATTTATTCCCACCATGTGATGGAGAAGATGGAATATGTTACTCATCCTCAGAAGCT
TCCAGGGGTGACAAGAGATGGCTTGAGCAAACACTTCTTCCACAAACCATCAAGAGCTTA
CACAAACCGGAACAAGAAAACGACGTAAATAAATCAAACCGATCACGACTCTGAGTTAAC
CGGATCATCAGAAACAGGTGGCACAACCGGCAAAACAAGACCGGTTATGATCAACGG
TCAACAAAGAGGATGCAAGAAGATATTAGTACTCTACACAACTTCGGCAAGAATCGTCG
ACCGGAGAAAACAAAATTGGGTGATGCATCAATATCATTTAGGGATTATGAGGAAGAGAG
AGAAGGAGAACTTGTGGTCTCCAAGATATTTTATCAGACACAACCAAGACAGTGTGTTAG
TAATACTAATTGGTCTGATCACCATGGTTCCAAGGACGTGATCGGAATTGGTGTGCGAGA
TGAGATTTCCAGCGTAGCTGCCACGTTGCAGAGTCTTGGCTCCGGTGACGTCGTTTCTAG
GGTTAATATGCATCCCATACAAGATCCTTTGATGAGGGGACAGCCGAAGCTTCAAAGGG
AAGAGAGAACCAGCATGTGTCTGGCAGGTGCGAGGAAGTACATGATGGGATCATAACATC
ATCAATGTCTATCTCATATGATTCATGATCATCATATAATCAACATCATCAAATCGGAGA
TAGAAGAGAATTTACATGTCTATCATATATCCCATGACCCCTACTATCACATCACAAACA
TGAGTCAATCTTCCATGTTACAAGTACTATGCCCTTTTCAGCGGCAGCAATTAAGGGGTCG
GTCGTCTGGTTCCGGGATTAGAAGACCTAATTATGGGTTGTACCACAGCTACGTGTACAGA
AGACAATAATCACAAATGATTAAATTCGCAGGAGCATTTCAGAAGCAAACCTCAGCGAAA
TGCAGAGTGGTTAACGTTTCCACAATTCTGGAACCAAGCCGAATCAGATGATCAAACCG
AAGATTTTAAACAGAACCAAAAGGAAGCAGAGAAAATCTTGCAAAAAGCTCTGCTTAGCTG
TTGATCAATGCCGGAATGCTGAGCTATGACTGACTAGTCTCTGCCATTTAACTTACAAT
ATCACCAGAGGTTGCGATGAATGTTGATTCGCTCAAAGGAGAGCGGCCGCTCTAGACAGG
CCTCGTACCG

>G1672 Amino Acid Sequence (conserved domain in AA coordinates: 41-194)

MKNSKCNLIDSKLEHHHLCGSKHCPGCGRMIQAATKPNWVGLPAGVKFDPDQELIEHL
EAKVKGKEENKKWSSSHPLIDEFIPTIDGEDGICYTHPQKLPGVTRDGLSKHFFHKPSRA
YTTGTRKRRKIIQTDHDSLTGSSETRWHKTGKTRPVMINGQQRGCKILVLYTNFGKNR
RPEKTNWVMHQYHLGINEEEREGLVVSIFYQTQPRQCVSNTNWSDDHSGSKDVIGIGVG
DEISSVAATLQSLGSGDVSVRVNMPHTRSFDEGTAEASKGRENQHVSGTCEVHDGIIT
SSMSSHMHIDHNNQHQQIGDRREFHMSSTYPMTPTTTSQHESIFHVTSTMPFQRQQLRG
RSSGSGLEDLIMGCTTATCTEDNNHK*

>G1677 (24..1037)

CAGTACTAATTCGTGTGTGTTAATGGTTCTAGTTATGGATGATGAAGAGAGTAACAACG
TTGAAAGATATGACGACGTCGTATTGCCAGGGTTTAGGTTCCATCCCACTGATGAAGAAC
TCGTAAGTTTCTACTTTGAAACGGAAGTTTACACAAATCTCTTCCCTTTGATCTCATCA
AGAAAGTCGACATTTACAAATACGATCCATGGGACCTCCAAAGCTTGACGCGATGGGGG
AAAAAGAGTGGTACTTTTATTGTCTTAGAGACAGGAAATACCGCAACAGCACAAAGACCTA
ACCGAGTAACCTGGAGGTGGCTTCTGGAAGCAACCGGAACAGACCGGCCTATATACTCAT
TGGACTCCACTCGATGCATCGGTTTGAAGAAATCACTTGTGTTCTACCGTGGTCGAGCTG
CTAAAGGAGTCAAAACCGATTGGATGATGCATGAATTTCTCTCTCTCTCTCTGACT
CTCATCACTCATCATATCCCAATTACAATAACAAGAAGCAACACCTTAACAATAACAACA
ACAGCAAGGAGCTTCTTCAAACGATGCTTGGGCGATATGTAGAATATTTAAGAAGACAA
ATGCAGTATCCTCACAAAGATCAATCCCAATCTTGGGTTTATCCAACGATTCTCTGACA
ACAATCAACAGTCACACAACAACACCGCAACTCTCTTAGCTTCATCAGACGTTCTCAGCC

ACATATCAACAAGACAAAACCTTTATTCTCTTCCAGTCAACGAACCCGCAAGCTTCACAG
AATCAGCTGCTTCTTACTTCGCGTCTCAGATGCTCGGAGTCACGTACAATACAGCCAGAA
ACAACGGAACAGGGGATGCTCTGTTTCTGAGAAACAATGGAACAGGGGATGCTCTGGTTC
TGAGCAACAATGAGAATACTACTTCAACAACCTTGACTGGAGGGTTGACTCATGAGGTTT
CGAATGTAAGATCAATGGTGATGGAGGAGACTACGGGGAGTGAGATGTCGGCGACGTCGT
ATTCCACTAACAATTAAGATCATAGTACTATTAACACTTGAATTAGTGTAGACGTTGATC
ATCGCTAATATGTATTAATTTTTCTTGTCTTACTATAAACGAAAAA

>G1677 Amino Acid Sequence (conserved domain in AA coordinates: 17-181)

MVLVMDDEESNNVERYDDVLPGRFHPDDEELVSFYLRKVLHKSPLPDLIKKVDIYKY
DPWDLPLKLAAMGEKEWYFYCPDRKRYNRSTRPNRVTTGGGFWKATGTDRIYSLDSTRCIG
LKKSLVFYRGRAAKGVKTDWMMHEFRLPSLSDSHSSYPNYYNNKKQHLNNNNNSKELPSN
DAWAICRIFKKTNVSSQRSIPQSWVYPTIPDNNQQSHNNTATLLASSDVLISHISTRQNF
IPSPVNEPASFTESAASYFASQMLGVTYNTARNNGTGDALFLRNNGTGDALVLSNNENNY
FNNLTGGLTHEVPNVRSMVMEETTGESEMSATSYSTNN*

>G1765 (139..966)

TCCTTCGCAAGACCTTCTCTTATATAAGGAAGTTCATTTTCAATTTGGAGAGGACACGCTG
ACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAAGAATGACTTGATTGGTGTCTAAA
GAGATCGATGTAGTGAAGATGAGTGGCGAAGGTAAGTAAAGGATCATGAAGAAGAA
AACGAAGCACCATTCTGGGTTTCAGGTTTCATCCGACGGATGAAGAGCTTTTAGGATAC
TATCTTCGAAGAAAAGTAGAGAACAAAACCATCAAACCTCGAAGTATCAAACAGATCGAT
ATCTATAAGTACGATCCTTGGGATCTTCCAAGAGTGAGCAGCGTCGGAGAAAAGGAGTGG
TACTTCTTCTGTCATGAGAGGTAGGAAATACAGGAATAGCGTTCGACCAAACCGAGTGACC
GGTTCAGGTTTCTGGAAGCCACTGATATTGATAAACCGGTTTACTCCAATCTTGACTGT
GTTGGTCTCAAGAAATCTCTGGTTTACTATCTTGGTTCAGCCGGTAAAGGCACCAAAACC
GATTGGATGATGCATGAATTCGCCCTCCCCTCCACCACGAAAACCGACTCTCCAGCTCAA
CAAGCAGAGGTATGGACACTTTGCAGAATCTTCAAACGAGTCACATCTCAAAGAAACCCA
ACCATCTTACCACCAAAACCGAAAACCGGTTATCACTTTAACCGACACTTGTCTAAGACC
AGCAGCTTAGATTCCGACCACACGAGCCACCGTACAGTAGATTCCATGTCCCACGAGCCG
CCGCTTCCACAGCCACAGAATCCTTATTTGGAACCAACATATAGTTGGTTTTAATCAACCG
ACATATACTGGTAATGATAATAACCTCCTGATGAGTTTCTGGAACGGCAACGGTGAGAT
TTCATAGGAGACTCAGCAAGTTGGGATGAACCTTAGATCTGTTATAGATGGCAACACTAAA
CCCTAGTAATAAAGTTTCTTTTTCAGCTTTGTACAAAAGATAAAACAAACGGCAACC
GCTCTAGACAGGCCTCGTACCGGGATCCTCTAGCTAGAGCTTTCGTTTCGTATCATCGGT
TTCGACAACGTTCTGT

>G1765 Amino Acid Sequence (conserved domain in AA coordinates: 20-140)

MSGEGNLGKDHEEENEAPLPGRFHPDDEELLGYLRRKVENKTIKLELIKQIDIIKYDP
WDLPRVSSVGEKEWYFFCMRGRKYRNSVRPNRVTTGGGFWKATGIDKPVYSNLDVGLKKS
LVYYLGSAGKGTCTDWMHEFRLPSTTKTDSAPAQAEVWTLCRIFKRVTSQRNPTILPPN
RKPVITLTDCTSKTSSLDSDHTSHRTVDSMSHEPPLPQPQNPYWNQHVGFNPQPTYTGND
NNLLMSFWNGNGDFIGDSASWDELRSVIDGNTKP*

>G1777 (97..1878)

CTCGTACTTTATCACCTCCGTCGTTCTATAATACTCTCTCCGTCAATCATATCATTGT
CGACAATTTTATTCTGATCAGTTTAAAAATGATCCATGGATGATAATTTAAGCGGCGAG
GAAGAAGATTACTATTACTCTCCGATCAGGAATCTCTCAACGGGATTGATAATGATGAA
TCCGTTTCGATACCTGTTTCTTCCCGATCAAATACTGTCAAGGTTATTACGAAGGAATCA
CTTTTGGCTGCACAGAGGGAGGATTGCGGAGAGTGATGGAATTGTTATCGGTTAAGGAG
CACCATGCTCGGACTCTTCTTATACATTACCGATGGGATGTGGAGAAGTTGTTTGCTGTT
CTTGTTGAGAAAGGGAAGATAGCTTGTCTTCTGGTCTGGTGTACACTTCTTGAAAAC
CAAAGTTGTGATTCTTCCGTTTCTGTTTCTTCTCGATGATGAGTTGTGATATCTGCGTA
GAGGATGTACCGGTTATCAGCTGACAAGGATGGACTGTGGCCATAGCTTTTGCAATAAC
TGTGAGACTGGGCATTTTACTGTAAAGATAAATGAAGGTCAGAGCAAAAGGATTATATGC
ATGGCTCATAGTGAATGCTATTTGTGATGAAGATGTTGTGAGGCTCTAGTTAGTAAA
AGCCAACAGATTAGCTGAGAAGTTTGATCGTTTCTTCTTGAGTCGTATATCGAAGAT
AACAAAATGTTGAGTGGTGTCCGAGTACTCTCATTGTGGGAATGCCATACGTGTTGAG
GATGACGAGCTCTGTGAGGTTGAATGCTCTTGTGGTTTGAGTTCTGTTTCAGTTGTTCA
TCTCAAGCTCACTCCCCTTGTCTTGTGTGATGTGGGAATATGGAGAAAGAAGTGCTTT
GATGAGTCCGAGACTGTTAATTGGATAACTGTTACACAAAGCCGTGTCCCAAATGTCAC

AAGCCTGTTGAAAAGAATGGTGGATGCAATCTCGTGACTTGTCTTTGTGACAATCTTTT
TGTTGGTTGTGTGGTGAAGCTACTGGAAGGGACCACACTTGGGCTAGAATCTCGGGTCAT
AGTTGTGGTCGGTTCCAAGAAGATAAAGAGAAACAAATGGAGAGAGCGAAAAGGGATCTC
AAGCGGTATATGCATTATCATAACCGATACAAAGCACATATCGACTCTCCAAGCTAGAG
GCTAAGCTTAGTAATAATATTAGTAAAAAGGTGTCTATTTCAGAAAAGAGGGAGTTACAA
CTTAAAGACTTCAGCTGGGCTACCAATGGACTCCATCGGTTATTTAGATCAAGACGAGTT
CTTTCATATTACACCTTTTCGCATTTTACATGTTTGGAGATGAGCTGTTTAAAGATGAG
ATGAGCTCTGAGAAAGAGAAATAAAACAAAATCTGTTTGGAGATCAGCAGCAGCAGCTT
GAGGCTAATGTTGAGAACTTTCTAAGTTCTTGGAGGAACCTTTTGATCAATTTGCTGAT
GATAAGGTCATGCAGATAAGGATTCAAGTCATCAATTTGTGAGTTGCGGTCGATACACTC
TGCGAAAATATGTATGAATGCATTGAGAATGACTTGTGGGTTCTCTGCAACTTGGCATC
CACAACATTACTCCATACAGATCAACCGGCATAGAACGAGCATCTGATTTTTATAGTTCC
CAGAATTTCAAGGAAGCTGTTGGTCAGTCTTCCGATTGTGGATGGACGTCAGGCTCGAT
CAAGCTTTGGAGTCAGGGAAGTCGGAAGACACAAGTTGCTCTTCCGGGAAGCGTGCTAGA
ATAGACGAAAGTTACAGAAACAGCCAAACACCTTACTAGATTTAACTTGCCAGCGGAA
GCCATTGAGCGGAAATGAACACTTATCCTTCTTCACTCCCAATAACACCTTTTGTGCC
AAATAAAGTGTGTTACCCGGATATTTATAGCTCTAAACCCAATCCCCTCTGCTTAATTTG
TCAGTGACCTTACCTAACCTCTTCA

>G1777 Amino Acid Sequence (domain in AA coordinates:124-247)
MDDNLSGEEEDYYSDDQESLNGIDNDESVSIPVSSRSNTVKVITKESLLAAQREDLRRV
MELLSVKEHHARTLLIHYRWDVEKLFPAVLVEKGKDSLFSAGVTLLENQSCDSSVSGSSS
MMSCDICVEDVPGYQLTRMDCGHSFCNNCWTGHFTVKINEGQSKRIICMAHKCNAICDED
VVRLVSKSQPDLAEFDRFLLESYIEDNKMVKWCPSTPHCGNAIRVEDDELCEVECSG
LQFCFSCSSQAHSPCSCVMWELWRKKCFDESETVNWITVHTKPCPKCHKPVEKNGGCNLV
TCLCRQSFCLWLCGEATGRDHTWARISGHSCGRFQEDKEKQMERAKRDLKRYMHYHNRYKA
HIDSSKLEAKLSNNISKVVISSEKRELQKDFSWATNGLHRLFRSRRVLSYSYPFAFYMF
GDELPKDEMSSEEREIKQNLFEQQQLEANVEKLSKFLEPPDQFADDKVMQIRIQVIN
LSVAVDTLCENMYECIENDLLGSLQLGIHNITPYRSNGIERASDFYSSQNSKEAVGQSSD
CGWTSRLDQALESKSEDTSKSSGKRARIDESYRNSQTTLLDLNLPAAIERK*
>G1793 (59..1783)

AGTGATTATTGATTAACCCAAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTCTAACAACCTGGCTTGGCTTTCCTCTTTACCGAACAACCTCTTCTTTGCCTCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGGACAACCTTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAACCGGACGAAAACCAATCCAACACCTAGTAGCTTACAA
CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACAACAGTAGCTATCATGAGCTTCAAGA
GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
TGTAGACAAAGCTTACCATCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAAGACATCGATGGACTGGTGCATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCCTTAAGTACTGGGGTCCTTC
AACTACTACTAATTTCCCATTAACAACCTACGAGAAAGAGTAGAGGAAATGAAGCACAT
GACGAGACAAGAGTTCTGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATACCAACATGGAAGATGGCAAGGATCGG
CCGAGTCGCCGGAACAAAGACCTTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTTTCCTATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCACTTCAACCTTACCCTCTAAGCATTCAACAACCATTAGAGCC
TTTTCTATCTCTTCAGAACAAATGACATCTCTCATTACAACAACAATGCTCAGGATTC
CTCCTCTTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTCAGCAACAGTCGAGCCAGAACTCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACACCGGTATTGG

TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAAAACAGATTA
CGATATGCCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTTGGACCAGTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGTCTTTCACTATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates: 179-255, 281-349)

MNSNNWLGFPLSPNSSLPPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGGDEGGEVPKV
ADFLGVSKPDENQSNHLVAYNDSYYFHTNSLMPVQSNVVAACDSNTPNNSSYHELQ
ESAHLNQLSLTSMGTTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTGRYEAHLWDNSCRREGQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGP
STTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFTSEEEAAEYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESSRKREAEMIALGSSSFQYGGGSSTGSGSTSSRLQLQPYPLSIQOPLE
PFLSLQNNDISHYNNNNAHDSSSFNHSYIQTQLHLHQQTNNYLQQQSSQNSQQLYNAYL
HSNPALLHGLVSTSIVDNNNNNGSSGSYNTAAFLGNHIGIGSSSTVGSTEEFPTVKTD
YDMPSSDGTGGYSGWTSES VQGSNPGGVFTMWNE*

>G180 (54..629)

GTAATTACGATCTACAACAAGTGACATCGTCGTCGACGACGATTCAAGAGAATATGAAC
TCCTCGTTCTCTTTGAAGAAACCAATGTCTTAACCTTTTCTCTTCTTCTTCTCTCTT
CTCTTTCTTCTCTCTTCTTCCCATTCACAACCTCTCTCTCCACTACTACTACTCATGCAC
CTCTAGGGTTTTCTAATAATCTTCAGGGTGGAGGACCCCTGGGATCAAAGGTGGTTAATG
ATGATCAGGAGAATTTTGGAGGTGGAACATAACAATGATGCTCATTCTAATTCTTGGTGA
GATCAAATAGTGAAGTGGAGATATGAAGAACAAAGTGAAGATAAGGAGGAAACTAAGAG
AGCCAAGATTCTGTTTCCAAACCAAAGCGATGTTGATGTTCTTGACGATGGCTACAAAT
GGCGTAAATATGGTCAGAAAGTCGTCAAGAACAGCCTTCACCCAGGAGTTATTACAGAT
GCACACACAACAACGTAGGGTGAAAAAGAGAGTGAGCGACTATCGGAAGATTGTAGAA
TGGTGATTACTACTTACGAAGGTCTGTCAACAACCATTCCTCTGTGACTCCACTTCTC
CTGACCATGATTGTCTCTCTTCTTTTAAACATCTCTTTCTATATATCTATATATAGACAG
TTATATGTGCACATATAGATGTGTGATATATTGCATATTTGATATTGCATGTGTTTTCA
AGAGTATGTCATCAGATGTTATGCATATATTCTTGACTTGTGCTTATAGTATACATATG
TAATAATATATATTGACATTGGTAGTTCATTTCTGTTCAAACAAAAAAAAAAAAAA

>G180 Amino Acid Sequence (domain in AA coordinates: 118-174)

MNFLVPFEETNVLTFSSSSSSSLSSPSFPIHNSSSTTTTHAPLGFNNLQGGGPLGSKV
VNDDQENFGGTTNNDHNSWWSNSGSGDMKNKVKIRKRLREPRFCFQTKSDVDVLDG
YKWRKYGQKVVKNLSLHPRSYRCHTHNCRVKRVERLSEDCRMVITTYEGRHNHIPSDDS
TSPDHDCLSSF*

>G192 (63..959)

CTTTTTTCTCTTCTCTCCTCAGAGATTGGAAGCTTTTTGTCTCCCTGAGTAACCAAATT
CAATGGCCGACGATTGGGATCTCCACGCCGTAGTCAGAGGCTGCTCAGCCGTAAGCTCAT
CAGCTACTACCACCGTATATCCCCGGCGTTTCATCTCACACAAACCTATATTCACCG
TCGGACGACAAAGTAATGCCGTCTCCTTCGGAGAGATTGAGATCTCTACACACCGTTCA
CACAAGAATCTGTCGTCTCTTCGTTTTCTGTATAAACTACCCAGAAGAACCTAGAAAGC
CACAGAACCAGAAACGTCCTCTTTCTCTCTCTGCTTCTCCGGTAGCGTCACTAGCAAAC
CCAGTGGCTCCAATACCTCTAGATCTAAAAGAAGAAAGATACAGCATAAGAAAGTGTGCC
ATGTAGCAGCAGAAGCTTTAAACTCCGATGTCTGGGCATGGCGAAAGTACGGACAGAAAC
CCATCAAAGGTTACCATATCCAAGAGGATACTACAGATGTAGTACATCAAAGGTTGTT
TAGCCCGTAAACAAGTGGAGCGAAATAGATCCGACCCGAAGATGTTTATCGTCACTTACA
CGGCGGAGCATAATCATCCAGCTCCGACACACCGTAATTCTCTCGCCGGAAGCACACGTC
AGAAACCATCCGATCAACAGACGAGTAAATCTCCGACGACCACTATTGCTACTTATTCAT
CGTCTCCGGTGACTTCAGCCGACGAATTTGTTTTGCCTGTTGAGGATCATCTAGCGGTGG
GAGATCTTGACGGAGAAGAAGATCTGTTATCTTTGTGCGATACGGTGGTTAGCGATGATT
TCTTCGATGGGTAGAGGAATTTCGACGCCGAGATAGCTTTTCCGGGAACCTCGGCTCCGG
CGAGTTTTGATCTCTCTTGGGTTGTGAACAGTGCCGCCACTACCACCGGAGGAATATGAT
TAGATTACGACGGCTTAGAATACTCTTATTAGGACAGATTATAGGATTAAGGAATTATT
CTCGGAGCATAGTAAAAATAGGATAAAAGAAATGTTCTTTGTTACTTTTTTTTCGGGTT
TTCTTCTTATTGTTTCTAAACATCTTAGAAAAAATTTAATTGTATATTCCTTAAGCTCGA
TACATCTTGTTTTAAAAAAAAAAAAAAAAAA

>G192 Amino Acid Sequence (domain in AA coordinates: 128-185)

MADDWDLHAVVRGCSAVSSSATTTVYSPGVSSHTNPIFTVGRQSNVSVFGEIRDLYTPFT
QESVVSFSCINYPPEPRKPQNQKRPLSLSASSGSVTSKPSGNTSRSKRRKIQHKVCH
VAAEALNSDVWAWRKYGQKPIKGSYPYRGYYRCSTSKGCLARKQVERNRSDPKMFIVTYT
AEHNHPAPTHRNSLAGSTRQKPSDQQTSSPTTTIATYSSSPVTSADFEVLPVEDHLAVG
DLDGEEEDLLSLSDTVVSDDFPDGLEEFAAGDSFSGNSAPASFDLSWVNSAATTTGGI*

>G1948 (18..1118)

AAAAGGTCTTCTTGCCATGGATACTTGTGCTCTAGTAATCCATCAGTCTCTGTCTCGCA
TCAAACCTTTCTCTCCCAAATCTTCTTCTTCTTCTTCTGCTTTCTCCCTGAATCCT
TACCGATCAGACGGATCGAGCTGTGTTTCCGAGGAGCTATATGTGCCGCCGTACAAAGAA
ACTACGAAGAAACGACCTCCTCCGTGGAAGAGGCAGAGGAAGATGATGAGTCATCATCAT
CGTACGGAGAAGTGAACAAGATCATTGGAAGCCGAACGGCGGGGAAGGAGCCATGGAGT
ACCTTATCGAGTGAAGGACGGCCATTCTCCGTGCTGGGTTCATCGAGCTACATCGCAG
CAGACGTAGTGTCCGAGTACGAGACACCCTGGTGGACGGCAGCTAGAAAAGCCGACGAGC
AGGCCCTGTACAGCTCCTGGAGGACCGAGACGTGATGCCGTGGACGAAAACGGCCGGA
CGGCTCTGCTTTTTCGTGGCAGGTCTGGGGTCCGACAAGTGCCTAAGGCTTCTGGCGGAGG
CTGGAGCCGATCTCGACCACCGAGACATGAGGGGAGGCTTGACGGCGCTGCACATGGCGG
CTGGTTACGTGAGGCCGGAGGTGGTGGAGGCGCTGGTGGAGCTGGGAGCTGATATTGAAG
TGGAAGACGAGAGAGGGTTAACGGCGTTGGAAC TAGCGAGGAGATTCTGAAGACGACGC
CGAAGGGGAATCCGATGCAGTTCCGGGAGGAGAATTGGGTTAGAGAAAGTGATCAATGTCC
TGGAAGGACAAGTGTTCCGAGTACGCCGAGGTGGATGAGATCGTAGAGAAACGAGGGAAAG
GCAAAGACGTTGAATATCTGGTCAGATGGAAGACGGTGGAGATTGCGAGTGGGTGAAAG
GTGTACACGTGGCGGAAGATGTGGCTAAGGACTACGAGGATGGGCTGGAGTACGCTGTAG
CGGAGAGTGTGATCGGGAAGAGGCTGGGAGACGATGGGAAGACCATCGAGTATCTTGTCA
AATGGACTGATATGTCTGATGCCACTTGGGAGCCTCAGGACAATGTCGACTCTACTCTTG
TTCTACTCTACCAACAACAACCAATGAATGAATGATTGATTTTGTGATTACATTCT
TCTCAATTTGCTTCTTTCTCATATGTGTTGGTTCATCTGACCGGTTCCGTTGGTACGTAC
CGGTACATTTTTCATTTTCTTTTAAGATGTGATCTTGATGGTTTTTGGCCTTTTGGGGACA
CTATTGATTTTATATCCATGCTTTGAATTTTGCTTCCCTTTTGGGGAGATTTCATGAAA

>G1948 Amino Acid Sequence (domain in AA coordinates: entire protein)

MDTCALVIHQSLSRILKSPKSSSSSSSAFSPESLPIRRIELCFRGAICAAVQRNYEETT
SSVEAEEDDESSSSSYGEVNIIGSRTAGEGAMEYLI EWKDGHSWVPSYIAADVSE
YETPWWTAAARKADEQALSQLEDRD VDAVDENGRTALLFVAGLGSDKCVRLLEAGADLD
HRDMRGGTLALHMAAGYVRPEVVEALVELGADIEVEDERGLTALELAREILKTPKGNFM
QFGRRIGLEKVINVLEGOVFEYAEVDEIVEKRGKGDVEYLVVRWKDGGDCEWVKGVHVAE
DVAKDYEGLLEYAVAESVIGKRVGDDGKTIEYLVKWTDMSDATWEPQDNVDSTLVLLYQQ
QQPMNE*

>G2123 (1..657)

ATGAGAAAAGTATGTGAGCTTGATATAGAGCTAAGTGAAGAGGAAAGAGACCTACTAACA
ACTGGATACAGAATGTCTATGGAGGCTAAGAGAGTTTCATTGAGAGTAATATCATCCATT
GAAAAAATGGAAGACTCGAAAGGAAACGACCAAAATGTGAAACTGATAAAAGGACAACAA
GAAATGGTTAAATATGAGTTTTTCAATGTTTGTAATGACATTTTGTCTCTCATTGATTCT
CATCTCATACCATCAACTACTACTAATGTGCAATCAATTGTCTTTTAAACAGAGTGAAA
GGAGATTATTTTCGATATATATGGCAGAGTTTGGTTCTGATGCTGAACGTAAAGAAAATGCA
GATAATTCTCTAGATGCATATAAGGTTGCAATGGAAATGGCAGAGAATAGTTTAGCACCC
ACCAATATGGTTAGACTTGGATTGGCTTTAAATTTCTCGATATTCAATTATGAGATCCAT
AAATCTATTGAAAGCGCATGTAAATTGGTTAAGAAAGCTTACGATGAAGCAATCACTGAA
CTCGATGGCCTTGACAAGAATATATGCGAAGAGAGCATGTATATCATAGAGATGCTTAAA
TACAATCTTTCTACGTGGACTTCAGGCGATGGTAATGGTAATAAGACAGACGGTTAG

>G2123 Amino Acid Sequence (domain in AA coordinates: 99-109)

MRKVCELDIELSEERDLTTGYKNVMEAKRVSLRVISSIEKMEDSKGNDQNVKLIKQQ
EMVKYEFNVNCDILSLIDSHLIPSTTTNVESIVLFNRVKGDYFRYMAEFGSDAERKENA
DNSLDAYKVAMEMAENSLAPTNMVRGLALNFSIFNYEIHKSIESACKLVKKAYDEAITE
LDGLDKNICEESMYIIEMLKYNLSTWTSBGDNGNKTDG*

>G2138 (27..512)

GAACCCCTAATTTCCGCAAAATTCATATGAAGCGTATTATCAGAATCTCATTACCCGACG
CAGAAGCCACCGATTCTTCTAGCGACGAAGACACGAGGAGCGTGGAGGAGCATCCCAGA
CTCGGCGCGGTGGGAAACGCCTCGTTAAAGAGATCGTAATCGATCCTTCCGATTCCGCCG

ATAAACCTCGATGCTGCTGCAAAACACACGGTTCAAATCAGGATCCCGGGCGGAATTTCTCAAGA
 CGCGGAAAAACGGAGAAGAAATATCGTGGAGTGAGGCAGAGGCCGTGGGGGAAGTGGGTGG
 CGGAGATCAGATGTGGAAGAGGAGCTTGTAAGGACGACGTGATCGTCTCTGGCTGGGTGTA
 CTTTAACTACTGCTGAGGAAGCTGCTCTAGCTTATGATAACGCTTCAATTAAGCTGATTG
 GACCTCACGCGCCGACCAATTTTGGTTTTGCCGGCGGAGAATCAAGAGGATAAGACGGTGA
 TTGGAGCTTCTGAGGTTGCTAGAGGCGCGTGAAGTGGGGTTGGTAATTTAGTTGTTAGC
 >G2138 Amino Acid Sequence (domain in AA coordinates: TBD)
 MKRIIRISFTDAEATGDSSEDETERGGASQTRRRGKRLVKEIVIDPDSADKLDVCKTR
 FKIRIPAEFLKTAKEKRYGRVQRQRPWKWVAEIRCGRGACKGRRDRLLWLTGTFNTAEAA
 LAYDNASIKLIGHPAPNFGFLPAENQEDKTVIGASEVARGA*
 >G2139 (40..663)
 CCTACAAGAAATCAAACACTAGTTCTGGTTTTCTGCAAACATGTCATCTACGAAGCAAGCA
 AAGGGAAGAAAAACAAGGGGAAGCAAAAGATCGAGATGAAGAAGGTGGAGAATATGGA
 GATAGGATGATTACGTTCTCAAACGCTAAAACCGGAATTTTAAGAAAATGAACGAGCTC
 GTAGCAATGTGTGACGTTCTGAACGGCTTTCTTGATTTTCTCTCAACCCAAGAAGCCCTAT
 ACATTGCGACATCCGCTCTATGAAGAAAGTGGCTGACCGGTTAAAGAACCCTTCGAGACAA
 GAACCATTAGAGAGAGACGATACCAGACCCCTCGTCGAAGCTTATAAGAAACGAAGGCTC
 CACGACCTCGTAAAAAAATGGAGGCGCTCGAAGAGGAGCTTGCGATGGATCTAGAGAAG
 TTGAAACTGTTGAAGGAATCGAGAAATGAAAAAGAAGTTAGATAAAATGTGGTGGAACTTT
 CCTTCGGAAGGTTTGTAGCGCGAAGGAGCTGCAGCAAAGGTACCAAGCGATGCTCGAGTTA
 CGTGATAACTTATGCGACAATATGGCTCACTTACGATTGGGAAAAGACTTGTGGTTCA
 TCTCTGTGTGTTGGGACGCTGAGTTTCTGGTGGTGTCTGCTGTGTTTCGATCGTGAAGCA
 TGATCATACATATTCTACTTGTATGATTAAATTTCTTTGTATTTGAAGTGTGATTTTA
 ATACTGCTATGATCCATTTGACGAAGCTCAATCGTCTCGAGTATATCTCTATTATCTAAC
 AGTATTGAGAAAAAAGGAGTTTCAGTAAAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
 >G2139 Amino Acid Sequence (conserved domain in AA coordinates:14-69)
 MSSTKQAKGRKTKGKQKIEMKKVENYGDRMITFSKRKTGIFKKMNELVAMCDVEVAFLIF
 SQPKPPLYTFAHPSMKKVNADRLKNPSRQEPLEDDTRPLVEAYKKRRLHDLVKKMEALEEE
 LAMDELKLLKESRNEKKLDKMWNFPSEGLSAKELQQRVQAMLELRDNLCDNMAHLRL
 GKDCGSSSVRVGRVSGVRLFDREA*
 >G2343 (1..1113)
 ATGGGTCATCACTCATGCTGCAACCAGCAAAAGGTGAAGAGAGGGCTTTGGTCAACGGAA
 GAAGATGAGAAGCTTATTAGATATATCACAACATCATGGCTATGGATGTTGGAGTGAAGTC
 CCTGAAAAAGCAGGGCTTCAAAGATGTGGAAAAAGTTGTAGATTGCGATGGATAAATAT
 CTTTCAGCTGATATCAGGAGAGGAAGGTTCTCTCCAGAAGAGAGAATGTATCATATAAGC
 CTTACGAGATTGTGGGAAACAGGTGGGCTCATATAGCTAGTCACTTTACCGGGAAGAACA
 GATAACGAGATTAAAAACTATTGGAATTCATGGATTAAGAAAAAGATACGAAAACCGCAC
 CATCATTACAGTCGTCATCAACCGTCAGTAACTACTGTGACATTGAATGCGGACACTACA
 TCGATTGCCACTACCATCGAGGCCTCTACCACCACAACATCGACTATCGATAACTTACAT
 TTTGACGGTTTCACTGATTCTCCTAACCAATTAAATTTACCAATGATCAAGAACTAAT
 ATAAAGATTCAAGAAATTTTTTCTCCATAAACCTCCTCTCTTCATGGTAGACACAACA
 CTTCTATCTTAGAAGGAATGTTCTTGAAACATCATCACAACAATAACAAGAACAAAT
 GATCATGATGACGCAAGAGGAGGGAAGAGAAAATGTTTGTGAACAAGCATTTCTAACA
 ACTAACACGGAAGAATGGGATATGAATCTTCGTCAGCAAGAGCCGTTTCAAGTTCCTACA
 CTGGCGTCACATGTGTTCAACAACCTCTTCCAATTCAAATATTGACACGGTTATAAGTTAT
 AATCTACCGGCGTAAATAGAGGGAAATGTCGATAACATCGTCCATAATGAAAACAGCAAT
 GTCCAAGATGGAGAAATGGCGTCCACATTCGAATGTTTAAAGAGGCAAGAATAAGCTAT
 GATCAATGGGACGATTCACAACAATGCTCTAACTTTTCTTTTGGGACAACCTTAATATA
 AACGTTGAAGTTTCACTCTTGTGTGGAAACCAAGCCCATCAATGAATTTGGGATCATCT
 GCCTTATCTTCTTCTTCCCTTCTCGTTTTAA
 >G2343 Amino Acid Sequence (domain in AA coordinates: 14-116)
 MGHHSNCCNQKVKRLWSPEEDEKLIRYITTHGYGCWSEVPEKAGLQRCGKSCRLRWINY
 LRPDIRRGRFSPEEBKLIISLHVGVGNRWAHIAHLPGRTDNEIKNYWNSWIKKIRKPH
 HHYSRHRQPSVTTVTNLADTTTSIATTIEASTTTTSTIDNLHFDGFTDSPNQLNFTNDQETN
 IKIQETFFSHKPPFLMVDTTLPILGEMFSENIITNNKNNDHDDTQRGGRNVEVCEQAFLT
 TNEEWDMLNRQEQBPVPTLASHVFNNSNSNIDTVISYNLPALIEGNVDNIHVNENSN
 VODGEMSTFECLKROELSYDOWDSDQCSNFFFDWNLINVEGSSSLVGNQDPSMNLGSS

ALSSSFPSF*

>G265 (280..1317)

CTTTGGTCTTGGAAGCCAAATCAAACCTTTCCTTCAATCCTCAAATTTTCGAAAATTTTC
TCTTTTGCTTTACGTTCTCTCAATTCTTATTTGTAAGAAAGTTTGTTCCTTTAATCAATC
AAATCAAAGAGACTTTTGAAGATTGTTTCCCAATTTGCGTCAATCGGGATCGAGTCAAAT
CTGAAATCTTCTCCACTCATCATCTGACTATAAGACTTAATCAAGGGACTTTTGTTCGG
GTTTGGTTTAAACGTCTTGGATTGGAAGTGGTTAAGGTATGGATGAAAATAATGGAGGT
TCAAGCTCACTCCACCTTTCCTTACTAAACATATGAAATGGTTGATGATTCTTCTTCT
GACTCGGTCTGTGCTTGGAGCGAAAACAACAAAGCTTCATCGTCAAGAATCCAGCAGAG
TTTTCAAGAGACCTTCTTCCGAGATTCTTCAAGCATAAGAATTCTCAAGTTTCATCCGT
CAGCTTAATACATATGGTTTTCGAAAAGTAGATCCTGAGAAATGGGAATTCTTGAATGAT
GATTTTGTAGAGGTGACCTTACCTTATGAAGAACATTTCATAGACGAAAACCGGTTTCAT
AGCCACTCGTTAGTGAATCTACAAGCGCAAAATCCTTTGACGGAATCAGAAAACGCGAGC
ATGGAGGATCAGATAGAAAGACTGAAAAATGAGAAAGAAGGCCCTTCTTGGCGAGTTACAG
AACCAAGAGCAAGAACGGAAGAAGTTTGAGCTGCAAGTAACGACATTGAAAGATCGGTTA
CAACATATGGAACAACATCAGAAATCAATAGTGGCATATGTTTACAGGTTTGGGAAAA
CCAGGACTTTCATAAACCTCGAAAACCATGAGAGAAGAAAAAGAAGATTTCAAGAGAAC
TCTCTTCCCTCCAAGCAGTTTACACATAGAACAGGTGCAAAAGTTAGAATCTTCGCTAACG
TTTTGGGAGAATCTTGTATCGGAATCATGCGAGAAGAGCGGTTTGCAGTCATCAAGCATG
GATCATGATGCAGCTGAGTCAAGTCTAAGTATTGGCGATACACGACCCAAATCATCGAAG
ATTGATATGAATCAGAGCCGCCGTTACCGTTACTGCGCTGCTCCAAAAACAGGCGTT
AACGATGACTTTTGGGAACAATGTTTGACAGAGAACCCTGGATCAACCGGCAACAAGAA
GTTCACTCAGAGAGAAGAGATGTCGGTAATGATAATAATGGTAATAAGATTGGAAATCAA
AGGACGTATTGGTGAATTCAGGGAATGTAAATAACATTACAGAGAAAGCTTCTTGACAT
GAATGAGGTTTTTTGAAAATAGTTTCTTTTGGTTCCACTGAGATTATTGTATGTGTTCA
TTATTTATTACTCTGTTTCTGTAAAAACAATCTCTCTATTGTTTGAGGCAGGAGTGACA
TAAATGCATATGCAGAATTGGTTTCAAAA

>G265 Amino Acid Sequence (domain in AA coordinates: 11-105)
MDENNGSSSLPPLTKTYEMVDDSSDSVVAWSENNKSFIVKNPAEFSRDLLPRFFKHK
NFSSFIRQLNTYGRKVDPEKWEFLNDDFVRGRPYLMKNIHRRKPVHSHSLVNLQAQNPL
TESERRSMEDQIERLKNEKEGLLAELQNEQERKEFELQVTTLKDRLQHMEQHOKSIVAY
VSQVLGKPGLSLNLNHHERRKRFRQENSLPPSSSHIEQVEKLESSLTFWENLVSECEKS
GLQSSSMHDHAESSLSIGDTRPKSSKIDMNSPEPVTVTAPAPKTGVNDDFWEQCLTENP
GSTEQEQVQSERRDVGNDNNGNKIGNQRTYWWNSGNVNNITEKAS*

>G2792 (1..960)

ATGGATCATCATCATCATAGCATCAAGAAATTCATCAACAACATCAGAATTACCATCA
TTTCAGCCAGCGTGCCATAACGGTAATGGTAACGGTTGGATCTATGACCCAAATCAAGTT
AGGTACGATCAAGTAGTGACCAACGGCTGTCAAAGTTGACGGATCTTGTAGGCAAGCAC
TGGTCAATTGCAACACCGAATAATCCCGACATGAACCATAACCTTCATCATCACTTCGAT
CATGATCATTCTCAAACGACGACATTTCTATGTACAGACAAGCCTTGAGGTTGAAAAAT
GAGGAAGATCTTTGTTACAATAATGGCTCAAGTGGTGGTGGTTCCTTGTTCATGATCCT
ATAGAAAGTTCTAGAAGTTTCTTGATATAAGGTTAAGTAGGCCATTAAACGGATATTAAT
CCGTCAATTAAAGCCATGCTTTAAGGCCTTAAACGTATCCGAGTTTAAACAAGAAAGAACAT
CAAACGGCATCTCTGGCAGCAGTGAGACTGGGAACAACAAACGCTGAAAAAAGAAGAGA
TGTGAAGAAATTTCCGATGAGGTTTCAAAGAAGGCCAAGTGCAAGTGAAGGCTCTACACTT
TCGCCAGAGAAGGAATACCCAAAGCCAACTTCGAGACAAGATCACGACTCTACAGCAA
ATTGTGTCTCCCTTTGGAAAGACTGATACTGCTTCTGTGCTTCAAGAGGCCATCACTTAC
ATAAATTTTATCAAGAGCAAGTTAAGCTGCTAAGCACTCCTTATATGAAGAATTCATCA
ATGAAGGATCCATGGGGGGGATGGGACAGAGAAGATCACAACAAAGGGGACCGAAGCAT
CTTGATCTAAGGAGTAGAGGGCTTTGTTTGGTTCCCTATTTTATATACCCCAATCGCATAC
CGCGATAACAGTGCAACTGACTACTGGAATCCACGTATAGAGGTTCTTTGTATCGTTAG
>G2792 Amino Acid Sequence (domain in AA coordinates: 190-258)
MDHHHHIASRNSSTTSELPSEFPACHNGNGNGWIYDPNQVRYDQSSDQRLSKLTDLVGKH
WSIAPPNNPDMNHNLHHFHDHDSQNDISMYRQALEVKNEEDLCYNNGSSGGGSLFHDP
IESSRSFLDIRLSRPLTDINPSFKPCFKALNVSEFNKKEHQASLAAVRLGTTNAGKKKR
CEEISDEVSKKAKCSEGSTLSPEKELPKAKLRDKITTLQQIVSPFGKTD TASVLQEAITY
INFYQEQVKLLSTPYMKNSSMKDPWGGWDREDHNKRGPKHLDLRSRGLCLVPISYTPAIY

>G2830 (1..903)

>G2830 Amino Acid Sequence (domain in AA coordinates:245-266)
MSSIPNRFNIYGGDTTNHRESLPIEMNHNHSMVRSMFITSDRMNHRDLFSSPPSFSSYQN
SHISSSSVGFNNSHMTYHMLKRNYSVSRADYFSTKDHSHFTQVSFTQTITNKYTTIVPS
NIFDTVHYDIGRVKRAIDFRNIWNPKSHLPKKFNRQCEILNPPLNIVFPHQDSADRQHL
DIIFSSSKHNHVFQDGRSLKKISEPTNLFEKSNSYDSQEDEKIDAYQYDGRTHSLPYTKY
GPYTCPRCNGVFDTSQKFAAHMLSHYNNETDKERDQRFRRNKKRYRKFMDSLKISKQKI
*

>G286 (94..2454)

TGCAATTTCCTCTCGACCAAAACCCTAATTTTCAGGTTTGGGGTTTTCTTCCTTTCACTGTC
AATTTTGATGAAACTTGTGATTCACTGATTAGAATGAATGCTAATGAGCAAACCTCGATCC
GCCAATGGCATTGGCAATGGCAATGGTGAGTCTATTCCCGGGATTCCAGATGACTTACGG
TGCAAGAGATCCGGATGGTAAACAGTGGAGATGCATGCAATGTCCATTGGCTGATAAGACT
GTTTGTGAGAGAAGCTACATCCAAGCAAGAAGCGGGCGGCTAATTCGCTTTTCAGGGCG
AACCAGAGAAGAACCAAAAGGCGATCATCGTTAGGCGAAACAGATACGTATTCGGAAGG
AAGATGGATGATTTCGAGTTACCACTCACCAGCATTGACCACTATAATAACGGTCTTGCC
TCTGCTTCCAAGAGTAATGGTAGACTAGAGAAGAGACATAATAAAAGCCTGATGCGGTAC
TCGCCCGAGACACCGATGATGAGGAGTPTCTCTCCACGTGTTGCAGTGGATTGAATGAT
GACTTGGGTAGAGATGTTGTAATGTTTGAAGAGGGCTACAGATCTTATAGGACACCACCA
TCTGTTGCTGTTTATGGATCCGACAGCAACAGATCACCAAGACACCGATCTTATGGAA
TACTCAGCAGCAAGCACAGATGTGCTGCAGAGTCTTTGGGGGAAACTGTGCCATCAATGC
CAGAGAAAAGATAGAGAGAGAATCATTTCTTGCCCTCAAATGCAATCAAAGAGCCTTCTGC
CACAATTGTCTATCGGCAAGGTACTCGGAGATATCACTTGAAGAAGTCGAGAAAGTTTGC
CCTGCATGTGCTGGCTTGTGTGATTGCAAATCTTGCTGCGTTAGATAATACAATAAAG
GTTCCGATCCGGGAAATACCCGTTTTTGGACAAGTTGCAGTATCTTTATCGTCTATTATCA
GCTGCTCTACCAGTCATAAAGCAGATCCATCTTGAAACAATGTATGGAAGTTGAACTAGAG
AAGAGGCTTTCGTGAAGTTGAGATTGATCTTGTCAGGCGAAGATTGAAAGCAGATGAGCAG
ATGTGCTGCACGTGTGTGCGATACCAGTTGTTGACTACTACCGTCACTGTCCGAACTGC
TCATATGACCTTTGCCCTGAGATGCTGTCAAGATCTACGGGAAGAGTCTTCAGTGACGATT
AGTGGGACTAACCAAAACGTACAAGATAGAAAAGGAGCTCCCAAACATAAACTTAACTTT
TCATACAAGTTTCTTGAGTGGAAGCCAACGGTGATGGGAGCATCCCTTGCCCTCCTAAG
GAGTATGGAGGCTGCGGTTACATCTTTGAATCTTGCCCGCATTTTCAAGATGAATTGG
GTTGCAAAGTATGTGAAAATGCTGAGGAGATTGTTAGTGCGTCAAATATCTGATCTT
CTGAACCTGATGTGTGTGATTCAAGATTCTGCAAATTTGCTGAGAGAGAAGAGAGCGGT
GACAACCTACGTGTACAGCCCGTCGCTTGAAACGATTAAAACTGATGGAGTAGCTAAGTTT
GAGCAACAATGGGCAGAGGGTCGGCTTGTTACTGTGAAAATGGTACTTGATGACTCATCT
TGCTCTAGATGGGATCCTGAGACTAATTTGGAGGGATATAGACGAGCTTTCGGACGAGAAA
CTGAGAGAACATGATCCATTCTTGAAGGCCAATTAATTGCTTGGATGGTTTAGAGGTTGAT
GTAAGACTTGGGGAGTTTACAAGAGCATATAAAGATGGAAGAACCAAGAGACAGGCTCTT
CCGCTATTGTGGAAGTTAAAGGACTGGCCGAGCCCAAGTGCTTCCGAGGAGTTCAATTTTC

TACCAAAGACCTGAGTTTATCAGAAGTTTTCCGTTTCTCGAGTACATTTCATCCCCGGTTA
GGCCTTCTGAATGTTGCAGCCAAGTTACCTCATTACTCGCTCCAAAACGATTTCAGGTCCA
AAGATTTATGTGCTTGTGGGACGTACCAAGAAATCAGTGCTGGCGATTTCATTGACTGGT
ATTCACTACAACATGCGTGACATGGTATACCTATTGGTGACACGTCTGAAGAAACAACA
TTCCGAAAGGGTGAGAAAAACAAACCTGTTCCAGAGGAACCTGACCAGAAGATGAGCGAA
AATGAGTCACTTCTTAGCCCTGAGCAGAAATTAAGGGACGGAGAGTTACATGATCTATCA
CTTGGTGAAGCCAGTATGGAGAAGAATGAACCTGAGTTGGCGTTGACTGTGAATCCAGAG
AACTTAACGGAAAAACGGTGACAACATGGAATCTTCTGCACATCTTCATGTGCAGGAGGA
GCCCAGTGGGATGTCTTTCGACGCCAAGACGTCCCAAAGTTGTCCGGGTATTTGCAGAGA
ACATTCAGAAAGCCTGATAATATCCAGACTGATTTTGAAGCCGTACCTGCTAATTCAAA
TAAATGAAGTGTGTAAGTCTTGTATGTGGAATGATTGAGTTTCTAGTTTGTACTCT
GGTTTCAGGTGTCAACCCGTTTGTATGAAGGATGTCTTTAAATGAACACCACAAGAGAC
AACTAAGAGACGAGTTTGGAGTTGAGCCATGGACATTTGAGCAACATCGTGGTGAGGCTA
TCTTCATTCCGGCTGGATGTCCGTTCCAAATCACTAATCTTCAGTCCAATATTCAGGTGG
CACTTGACTTCTTGTGCCCTGAAAGCGTTGGAGAGTCAGCAAGACTAGCTGAAGAAATCC
GGTGTTTACCAAACGACCACGAGGCAAAACTTCAGATTCTAGAGATTGGAAGATATCAT
TATACGCAGCTAGCTCAGCCATTAAAGAGGTTTCAGAACTGGTCTTGGATCCAAAGTTTG
GAGCAGAGCTTGGATTTGAAGACTCTAACTTAACCAAAGCAGTCTCTCACAACCTAGACG
AGGCAACCAAGCGGCC

>G286 Amino Acid Sequence (domain in AA coordinates: TBD)
MNANEQTRSANGIGNNGESIPGIPDDLRCRSDGKQWRCTAMSMADKTVEKHYIQAKK
RAANSAFRANQKAKRRSSLGETDITYSEGKMDDFELPVTSIDHYNNGLASASKSNGRLEK
RHNKSLMRYSPETPMRFSFPRVAVDLNDDLGRDVMFEEGYRSYRTPPSVAVMDPTRNR
SHQSTSPMEYSAASTDVSAESLGEICHQCQQRDRERIISCLKCNQRAFCHNCLSARYSEI
SLEEVEKVCPCACGLCDCKSLRSDNTIKVRIREIPVLDKLYRLLSAVLPVIKQIHL
EQCMEVELEKRLLEVEIDLVRARLKADEQMCCNVCRIPVVDYYRHCPNCSYDLCLRCCQD
LREESSVTISGTNQNVQDRKGAPKLKLNFSYKFPWEANGDGSIPCPPKEYGGCGSHSLN
LARIKMNWVAKLVKNABEIVSGCKLSDLLNPDMDCSRFCFAEREESGDNYVYSPSLET
IKTDGVAKFEQQWAEGRLLVTKMVLDDSSCSRWDPETIWRDIDELSDEKLREHDPFLKAI
NCLDGLLEVDRLGEFTRAYKDGNQETGLPILLWKLKDWSPSPASEEFIYQRPFIERSFP
FLEYIHPRLGLLNVAALKPHYSLNDSGPKIYVSCGTYQEISAGDSLTIHYNMRDMVYL
LVHTSEETTFEVRVKTTPVPEEPDQKMSENESLLSPEQKLKRDGELHDLISLGEASMEKNEP
ELALTVPENLNTENGDNMESSCTSSCAGGAQWDVFRQDVPKLSGYLQRTFQKPDNIQTD
FVSRTC*

>G291 (124..1197)
CAAGAACCCAAAGATCTCTCTCTATTGTTTGCCTTCTTCTTTCTTTCTGACTCAAACCC
TCAAATCAATTCTCGCGATTAAGCAAAACCTAGATTATTCTACTCTTCGAAGTCGATT
TCAATGGAAAGGTTCTCTCGTCAGCCATCGCGAGGAAGACATGGGAGCTAGAGAACAACATT
CTCCAGTGAACCAACCGATTGAGCTCCGACAGTATATTCCACTACGACGACGCTTCA
CAAGCCAAAATCCAGCAGGAGAAGCCATGGGCCTCCGATCCTAACTACTTCAAGCGCGTT
CACATCTCAGCCCTTGCTCTTCTCAAGATGGTGGTTACGCTCGCTCCGGTGGCACAATC
GAGATCATGGGTCTTATGCAGGGTAAAACCGAGGTGATACAATCATCGTTATGGATGCT
TTTGCTTTGCTGTTGAAGGTACTGAGACTAGGGTTAATGCTCAGTCTGATGCCATGAG
TATATGGTTGAATACTCTCAGACCAAGCAAGCTGGCTGGGAGGTTGAGAACGTTGTTGGA
TGGTATCACTCTCACCCTGGGTATGGATGTTGGCTCTCGGGTATTGATGTTTCGACACAG
ATGCTTAACCAACAGTATCAGGAGCCATTCTTAGCTGTTGTTATTGATCCAACAAGGACT
GTTTCGGCTGGTAAGGTTGAGATTGGGGCATTGAGAACATATCCAGAGGGACATAAGATC
TCGGATGATCATGTTCTGAGTATCAGACTATCCCTCTTAACAAGATTGAGGACTTTGGT
GTACATTGCAAAACAGTACTACTCATTTGGACATCACTTATTTCAAGTCATCTCTCGATAGT
CACCTTCTGGATCTCTTTTGAACAAGTACTGGGTGAACACTCTTCTTCTTCCCACTG
TTGGGCAATGGAGACTATGTTGCCGGGCAAATATCAGACTTGGCTGAGAAGCTCGAGCAA
GCGGAGAGTCAGCTCGCTAACTCCCGGTATGGAGGAATTGCGCCAGCCGGTCACCAAAGG
AGGAAAGAGGATGAGCCTCAACTCGCGAAGATAACTCGGGATAGTGCAAAGATAACTGTC
GAGCAGGTCCATGGACTAATGTACAGGTTATCAAAGACATCTTGTTCAAATCCGCTCGT
CAGTCCAAGAGTCTGCTGACGACTCATCAGATCCAGAGCCCATGATTACATCGTGAAGT
TGGTCTATTCTTTTCTTTTGGCTGCGGAAATTGACTATCGGTTTGACCCGGTTTATGA
GGCAATGCCCATGTTTCCCTATATCTCTAGTGTAGTATCTGCTTCAGACAAAGATCTTTG

GGTTATTAAATGACATTAACATAAAAAAA
>G291 Amino Acid Sequence (domain in AA coordinates: 132-160)
MEGSSSAIARKTWELENNILPVEPTDSASDSIFHYDDASQAKIQQEKPWASDPNYFKRVH
ISALALLKMMVHARSGGTIEIMGLMQGKTEGDTIIVMDAFALPVEGTETRVNAQSDAYEY
MVEYSQTSKLAGRLNVVGVYHSHPGYGCWLSGIDVSTQMLNQYQEPFLAVVIDPRTV
SAGKVEIGAFRTYPEGHKISDDHVSEYQTIPLNKIEDFGVHCKQYYSLDITYFKSSLDSE
LLDLLWNKYWVNTLSSSPLLNGNDYVAGQISDLAEKLEQAESQLANSRYGGIAPAGHQRR
KEDEPQLAKITRDSAKITVEQVHGLMSQVIKDILFNSARQSKSADSSDPEPMITS*

>G427 (49..1230)

TTTCCTCTCCGAAACAGAAATTCAAAAACAAATTCACACGAAAACGATGGCGTTTCAT
AACATCACTTTAATCATTTTACCGACCAACAACAACATCAGCCTCCTCCTCCGCCGCAA
CAGCAGCAGCAACAACATTTTCAAGAATCAGCACCCCCCTAATTGGCTCCTCCGCTCCGAC
AACAACTTCTCAATCTCCACACAGCTGCCACAGCCGCCGCTACAAGCTCCGATTCTCCT
TCTTCCGCCGCCGCTAACCAGTGGCTCTCAGATCCTCATCCTTCTCCAACGAGGCAAC
ACCGCAAACAACAACAACAACGAAACATCCGGTGACGTATCGAAGACGTTCCCGCGCGGA
GAGGAGTCAATGATCGGAGAGAAGAAGGAGGCGGAGAGGTGGCAGAATGCGAGACACAAG
GCGGAGATACTGTCTCATCCACTATACGAGCAACTTTGTGCGCACACGTGGCGTGCCTG
AGGATCGCAACGCCGGTGGATCAGCTTCCGAGGATAGACGCACAGCTTGCTCAGTCTCAA
AAGCTCGTGGCTAAGTACTCAACTTTAGAAGCCGCTCAAGGACTCCTCGCCGCGCATGAC
AAGGAGCTTGACCACTTCATGACGCATTATGTACTATTGCTTTGCTCTTCAAAGAACAA
CTGCAACAGCATGTTCTGTGTTTATGCAATGGAAGCTGTTATGGCCTGTTGGGAGATTGAA
CAGTCGCTTCAAAGTTTTACAGGAGTATCTCCTGGTGAAGGCACAGGAGCAACAATGTCT
GAGGATGAAGATGAGCAAGTAGAGAGTGTATGCTCATTTGTTTGTATGGAAGCTTAGATGGG
TTAGGGTTTGGTCTCTAGTTCCCACTGAGAGCGAGAGATCTTTGATGGAACGAGTCAGA
CAAGAACTCAAACATGAACCTCAAGCAGGGTTACAAGGAGAAAATTGTGGACATAAGAGAG
GAGATACTGAGGAAGAGAAGAGCTGGAAAATTACCAGGAGACACCACCTCTGTTCTCAA
TCATGGTGGCAATCTCATTCTAAGTGGCCTTACCCTACTGAGGAAGATAAGGCGAGGTTG
GTGCAGGAGACGGGTTTGCAGCTCAAACAGATAAACAATTGGTTCATCAATCAAAGAAAG
AGGAATTGGCATAGCAATCCATCTTCTTCTACCGTCTCAAAGAATAAACGCCGAAGCAAT
GCAGGTGAAAACAGCGGAAGAGACCGTTGAGATCAAGCTTGATGTAGAGATCCAAAAGC
TTTATAGAAAGGTGGAGGCATGAAGACAAGAATTCTTACACAACAACGTTAGGACGTAA
TTTTGTGCCAGTACATGGTATGGCTTTTCAATTTGGTAATGATTAGGGCCACACAAAATT
AAACCCCAAAGCATGATTTGTAATATGAGGTTTATGATGGACTTTATGATAGGATCGTCA
GTCTTCACTGCCATCTCCATTCTCCACCATCAATCCATCATTATATCTTGTGAAAAAAA
A

>G427 Amino Acid Sequence (domain in AA coordinates: 307-370)
MAFHNNHFNHFTDQQQHQP PPPPQQQQQHFQESAPPNWLLRSDNNFLNLHTAATAAATS
SDSPSSAAANQWLSRSSSFLQRGNTANNNNNETSGDVIEDVPGGEESMIGEKKEAERWQN
ARHKAIEILSHPLYEQLLSAHVACLRIATPVDQLPRIDAQLAQSONVVAKYSTLEAAQGLL
AGDDKELDHFMTHYVLLLCSPKEQLQHVVRVHAMEAVMACWEIEQSLQSFTGVSPGEGTG
ATMSEDEDEQVESDAHLFDGSLDGLGFGPLVPTESERSLMERVRQELKHELKQGYKEKIV
DIREILRKRKAGKLPDGTTSVLKSWWQSHSKWPTYTEEDKARLVQETGLQLKQINNFI
NQRKRWNHNSNPSSSTVSKNRRSNAGENSGRDR*

>G509 (122..1054)

CTTCCTCCTTTGCTAATAAACTTTTCTTTGAACCTTACACGCCTTGTGATATTACTCTC
TTAAATATATATTTTCGTACATTAACACAGACATATATAAGCTAAAGATTTCTTCACGT
AATGGGTTTGAAGATATTGGGTCCAAATTGCCACCGGGTTTCGATTTTCATCCAAGTGA
TGAAGAGTTGGTTTGTCTTATCTTTGCAACAAGATTAGGGCCAAATCTGATCATGGTGA
TGTTGATGATGATGATGATGATGTTGATGAAGCTTTGAAGGGTCTACTGATCTTGTGGA
GATTGACTTGATATCTGTGAGCCATGGGAGCTTCTGATGTGGCAAAGTTAAACGCAAA
GGAATGGTACTTCTTCAGTTTCCGTGATCGAAAGTATGCTACTGGATATCGCACGAACAG
AGCGACAGTAAGCGGATACTGGAAAGCAACAGGAAAAGATCGAACGGTGATGGATCCACG
TACAAGGCAATTGGTAGGGATGAGAAAAACTAGTGTTCTACAGAAACAGAGCACCAAA
TGGGATCAAAACTACTTGGATCATGCACGAGTTCCGTCTTGAGTGTCTTAACATCCCACA
TAAGGAAGACTGGGTCTTGTGCAGAGTGTTCACAAAGGCAGAGACTCATCGCTACAAGA
CAATAATTATTATAACAATGATAATCAGACGCAAGGCTTGAAGTTAATGACGCTCCGGA
TCTTAATTACAACAATCAGTTGCCACCTTTGCTATCATCCCTCCTCATATCATCAACA

TGAGAAGATGAAAATCCAAGTTTGTGATCAGTGGGAGCAGCTAATGAAGCAGCCTTCAAG
GACCACCGGCCACCCCTATCATCACCATTGTGCATCATCAAACCATAGCATGTGGTTGGGA
GCAGATGATGATCGGTTGCTGTGCATCACCTTCGAGTCATGGCCCTGATCAGAGTCCTT
TGCTAAATTTGCTTTACCGTCGACAATAACAACAGTGTCAACATCAGTGGTGATCATCAT
CAGAATTATGAGAAGATTTTGTGTGCATCACTAGACATGACGAGTTTGGATCAGGACAAG
ACATGTATGGGATCATCATCGGATGGTGGTATGGTCTCTGATCTTCACATGGAATGTGGT
GGATTGAGTTTGTAGACCGAAAATATCCTCGCTTTCCAATGAACATAATTCAAGGGGTTT
GCCAATTTGTGTGATTCGTGAATTATACAAACATTTTATCTATAGATTTATCACATTATCA
AACATGTAAGTTGTGTGGCATTGGGTATAGGGTTTGTGTGATTCTAGGTTTTTAGGACG
ATGTATGTTGTATATTAGCGTGTTTTAGGATTATTCTCATTTAAAATTATATGAA
AACCATTACTATGAATACAATTAGTTTCTTTGTTGTAAATAATTTTTAGATTATCAA
AAAAAAAAAAAAAA

>G509 Amino Acid Sequence (domain in AA coordinates: 13-169)
MGLKDIGSKLPPGFRFHPSPDEELVCHYLCNKIRAKSDHGDVDDDDDDVDEALKGSTDLVE
IDLHICEPWELPDVAKLNKEWYFFSFRDRKYATGYRTNRATVSGYWKATGKDRTVMDPR
TRQLVGMKRKTLVFNRRAPNGIKTWMHEFRLECPNIPHKEDWVLCRVFNKGRDSSLQD
NNYNNNDNQTRLEVDAPDLNYYNQLPPLLSSPPHNQHEKMKIQVCDQWEQLMKQPSR
TTGHPYHHCHHQTACGWEQMMIGSLSSPSSHGPDHESFAKFALPSTITTVSTSVVIII
RIMRRFCCHH*

>G519 (85..894)
CACAAAGATCCTCCGATTTCGAAGGTTTATAAAAACTCAAATCGAATCTTATCCACAAGA
AAACAACAAAGGTACTTTTCAAAAATGAAGGCGGAGTTGAATTTGCCGGCGGGATTCCGA
TTTCATCCGACGGACGAAGAGCTTGTCAAGTCTTATCTTTGCCGGAGATGTGCGTCAGAA
CCGATTAACGTTCCGGTTATCGCAGAGATTGACTTGTACAAATCAATCCATGGGAGCTT
CCAGAAATGGCGTTGTACGGTGAGAAAGAATGGTACTTCTTCTCGCATAGAGACCGGAAA
TACCCAAACGGGTCGAGACCAAACCGGGCAGCTGGAACCGGTTATTGGAAAGCGACTGGA
GCTGATAAACCGATCGGAAAACCGAAGACGTTAGGGATTAAGAAAGCACTCGTCTTCTAC
GCAGGAAAAGCTCCGAAAGGATTAAACGAATTGGATTATGCACGAGTATCGTCTCGCT
AATGTCGATCGATCTGCTTCTACCAACAAGAACAACAACTTAAGACTTGATGATTGGGTT
TTGTGTCGGATATACAATAAGAAAGGAACAATGGAGAAGTATTTACCGCGCGCGCTGAG
AAACCGACGAAAAGATGAGTACGTCCGACTCAAGATGCTCAAGTCACGTGATTTACCG
GACGTACGTTGTTCTGATAACTGGGAGGTTGAGAGTGAGCCCAAATGGATTAATCTGGAA
GACGCGTTAGAGGCATTTAATGATGACACGTCCATGTTTAGTTCCATTGGTTTGTGCAA
AATGACGCTTTGTTCTCAGTTTCACTACAGTCCCTCCGATTTCGTGATTCGTTTCTAG
GACCGTTTCGAGCAGAAACCGTTCTTGAATTGGAATTTTGCTCCTCAAGGGTAAAAATAA
TCGGCAAAAAGTTGAAGCTTTTCAGAGTCTTCGATCACCAGCATTGTGTGCGATCCTGAC
CCGGAGACCAAGTCGGGTACATACGATTACATAATCGGGTTATTGAGATTTCACATTGG
ATTTCCGAGACTAACCACTTAACGGATTCTGGGGTAATTGGGGGGTTTGCACAGGTGA
ATCACACTGAGTCAGCAAGTTTCGATTTTGTGTTTGTGTTTGTAAATGATTGATTAAATG
TCTAAAGATATACGAAGTAGATTGAGAAGCACTGTAAGCAATTGTGACCACCCGTTA
TGAATCATAAATATATTCAATGAAGCATGAGCTTATTTTAAAAA

>G519 Amino Acid Sequence (conserved domain in AA coordinates: 11-104)
MKAELNLPAGFRFHPDTEELVKFYLCRRCASEPINVPVIAIDLYKFNPWELPEMALYGE
KEWYFFSHRDRKYPNGSRPNRAAGTGYWKATGADKPIGPKTLGIKKALVIFYAGKAPKGI
KTNWIMHEYRLANVDRSASTNKKNNLRLLDDWVLCRIYNNKGTMEKYLPAAEKPTKEMST
SDSRCSHVISPVTCSNWEVESEPKWINLEDALEAFNDDTSMFSSIGLLQNDAFVPPQF
QYQSSDFVDSFQDPFEQKPLNWNFAPQG*

>G561 (86..1168)
AATTTGTTTTTTTTCTTTTGTGGGTTCAATTGCAATTGTTTTCCCTGAGACTCAAGTTA
CTGTGTCATTACTCTGCATTGAGCAATGGGTAGCAACGAAGAAGGAAACCCCACTAACAA
CTTGATAAGCCATCGCAAGCTGCTGCTCCTGAGCAGAGTAATGTTTCATGTGTATCATCA
TGACTGGGCTGCTATGCAGGCATATTATGGGCCTAGAGTTGGTATACCTCAATATTACAA
CTCAAATTTGGCGCCTGGTCATGCTCCACCGCTTATATGTGGGCGTCTCCATCGCCAAT
GATGGCTCCTTATGGAGCACCATATCCACCATTTTGCCCTCCTGGTGGAGTTTATGCTCA
TCCTGGTGTCAATAGGCTCACAACCACAAGGTCCTGTTTCTCAATCAGCATCTGGAGT
TACAACCCCTTTGACCATGTATGACACAGCTAATTCAGCTGGAACCTCAGATCATGGGTT
CATGAAAAGCTGAAAGAGTTTCGATGGACTTGCAATGTCAATAAGCAATAACAAAGTTGG

GAGTGTCTGAACATAGCAGCAGTGAACATAGGAGTTCTCAGAGCTCCGAGAATGATGGCTC
TAGCAATGGTAGTGATGGTAATAACAACCTGGGGGAGAACAATCTAGGAGGAAAAGAAGGCA
ACAAAGATCACCAAGCACTGGTGAAAGACCCTCATCTCAAAACAGTCTGCCTCTTAGAGG
TGAAATGAGAAACCCGATGTGACTATGGGGACTCCTGTTATGCCACAGCAATGAGTTT
CCAAAACCTCTGCTGGCATGAACGGTGTGCCACAGCCATGGAATGAAAAAGAGGTTAAACG
AGAGAAGAGAAAAACAGTCAAACCGAGAATCTGCTAGGAGGTCAAGACTGAGGAAGCAGGC
TGAAACAGAACAACTATCTGTCAAAGTTGACGCATTAGTAGCTGAGAACATGTCTCTGAG
GTC'TAAACTAGGCCAGCTAAACAATGAGTCTGAGAACTACGGCTGGAGAACGAAGCTAT
ATTGGATCAACTGAAAGCGCAAGCAACAGGGAAAACAGAGAACCTGATCTCTCGAGTTGA
TAAGAACAACCTCTGTATCAGGTAGCAAAACTGTGCAGCATCAACTGTTAAATGCAAGTCC
GATAACCGATCCTGTGCGGGCTAGCTGACCGTGCCGCAACAATGAGAACCCGATATTTT
TTCTTTGGGTTGTGATTGTAAC'TAAAAGGAGACTTTTTGTTTTTATTCTTAGATTTGT
AGCTCTCTGCATAGTGAGCATAAATTGATGTAATATGGTTTAAAGAGATTGGTGTCTCT
GGTGTGTGCTGCAACCACATAATTGGTGATAGATAGGTTTAGTTATATAAGCAAATGTAT
TAGAGATAAGGGGAGACATATTTGATGGTCTTT

>G561 Amino Acid Sequence (domain in AA coordinates: 248-308)
MGSNEEGNPTNNSDKPSQAAPEQSNVHVYHHDWAAMQAYYGPRVGIPOYYNSNLAPGHA
PPPYMWASPSPMAPYGAOPYPPFCPPGGVYAHPGVQMGSPQGPVSQSASGV'TPLTIDA
PANSAGNSDHGFMKKLKEFDGLAMSISNNKVGSAEHSSEHRSSQSSSENDGSSNGSDGNT
TGGEQSRKRQRQSPSTGERPSSQNSLPLRGENEKPDVTMGTPVMPTAMSFQNSAGMNG
VPQPWNEKEVKREKRQSNRESARRSRLRKAETEQLSVKVDALVAENMSLRSKLGQLNN
ESEKLRLENEAILDQLKAQATGKTENLISRVDKNNSVSGSKTVQHQLLNASPIITDPVAAS

*

>G590 (102..1223)
TCGACAGACACTCTCCCTCTCTCCATGCCCATAAAATCTCAAAGACTGTTTAAAAA
AATGTTTTAGCTTTAACTGCTTTTTTTTTTGTGTTGTTGGTGTAATGATATCAGAGAGAAG
AAAGAGAAGAGAAGAAGCAGAGAGTATGGGAGATAAGAAATTGATTTCATCTTCTTCTT
CTTCTCGGTTTACGATACTCGTATCAATCATCATCTTCATCATCCTCCGTCTTCTTCCG
ACGAAATCTCTCAGTTTCTCCGGCATATTTTCGACCGTTCTTCTCCTTTACCTTCTTACT
ACTCCCGGCGCAGACTACAACGACGGCGTCTTTGATTGGTGTGCACGGGAGCGGTGACC
CACATCGAGATACTCGAGAAGTCTCGTTTCTCATCATCCACCGTCAGATTCTGTGCTTA
TGTCGAAACGTGTGCGGAGATTTCTCTGAGGTTTAAATCGGCGGAGGATCAGGCTCAGCCG
CCGCGTGT'TTGGTTTCTCCGGTGGTGGTAATAATAACAACGTTCAAGGAAATAGCTCTG
GGACTCGAGTATCGTCTTCTTCCGTGGAGCTAGTGGCAACGAGACAGATGAGTATGACT
GTGAAAGCGAGGAAGGAGGAGAAGCTGTAGTTGATGAAGCTCCCTCTTCCAAGTCAGGTC
CTTCTTCTCGTAGTTTATCTAAAAGATGCAGAGCTGCTGAAGTTCATAATCTCTCTGAGA
AGAGGAGGAGAAGTAGAATTAATGAAAAATGAAAGCTTTACAAAGTCTCATCCCTAATT
CAAATAAGACCGGATAAGGCTTCAATGCTTGATGAAGCCATTGAGTATCTGAAACAGCTTC
AGCTCCAAGTTTCAGATGTTGACTATGAGAAATGGAATAAACTTGCATCCTTTGTGTTTAC
CTGGAACCTACATTACACCCATTGCAACTCTCTCAGATTGACCCCTGAAGCAACCAATG
ATCCTCTGCTTAATCATACCAATCAGTTTGCTTCGACTTCTAATGCACCGGAAATGATCA
ATACTGTGGCTTCTTCATACGCTTTGGAACCTTCTATTGCGAGTCACTTTGGACCTTTCC
CTCTCCTTACTTACCCGTGGAGATGAGTCGGGAAGGTGGGTAACTCATCCAAGGTTGA
ACATTGGTCAATCCAACGCAACATAACCGGGAACAAGCTCTGTTTGATGGACAACCTG
ACCTAAAAGATCGAATTACTTGAACAGTGTCCCAACTTCGGGATCTCTATGTGTTCTGT
TTCTTAGAACGCAAGCCATAAAGCTGTCTGAC

>G590 Amino Acid Sequence (domain in AA coordinates: 202-254)
MISQREEREKKQVRMGDKKLISSSSSSVYDTRINHLHHPSSSDEISQFLRHIFDRS
SPLPSYYSPTATTTTASLIGVHGSDPHADNSRSLVSHHPPSDSVLMSKRVGDFSEVLIG
GGSGSAAACFGFSGGGNNNVQGNSSGTRVSSSVGASGNETDEYDCSEEGGEAVVDEA
PSSKSGPSSRSSSKRCRAAEVHNLSKRRRSRINEKMKALQSLIPNSNKTDKASMLDEAI
EYLKQLQLQVQMLTMRNGINLHPLCLPGTTLHPLQLSQIREPEATNDPLLNHTNQFASTS
NAPEMINTVASSYALEPSIRSHFGPFPLLTSPVEMSREGGLTHPRLNIGHSNANITGEQA
LFDGQPDLDKRIT*

>G818 (65..1060)

GTATTTCTTACAATAAACGACCAAAAAGTTAATACAAGAAATAGAAACGGTGTAGGAAGC
TACTATGACGGCAATTCAAACGTCGTCGATATTGAATCTTCTTCTCTTCTCGCTTTGTCA

AGAGACGGCAACGGAGACCGTCACCGTTGAAAGAGGCTCGTCTGATTTCATCTTCAAAGCC
AGACGACGTCGTTTTACTAATCAAGGAAGAGGAGGATGACGCCGTTAACTTGTCACCTGG
TTTTTGAAAATTGCACGAGATAGGTTTTAATAACACCGTTCTTGAGAAAAGACGTTTGAGAT
CGTCGATGACAAAGTAACAGACCCGGTTGTATCATGGAGCCCGACCCGTAAAAGCTTTAT
CATTTGGGATTCTTACGAGTTCTCAGAGAATCTACTTCCCAAATACTTCAAGCACAAGAA
CTTCTCCAGTTTATTCTCGTCAGCTTAACTCTTACGGTTTTAAAAAGGTCGATTTCAGATAG
GTGGGAATTTGCTAACGAAGGTTTTCAAGGAGGGAAGAAACATTTGCTTAAGAATCAAA
GAGGAGAAGCAAAAACACTAAATGTTGTAACAAGGAAGCGAGTACCACCACGACAGAGAC
TGAGGTTGAGTCATTGAAGGAGGAACAGAGTCCAATGAGATTGGAGATGTTGAAGCTGAA
ACAACAACAGAAGAATCTCAACATCAGATGGTCACTGTGCAGGAGAAGATCCACGGAGT
TGATACCGAACAACAGCATATGCTTAGTTTCTTTGCAAAGTTGGCTAAAGATCAAAGATT
TGTAAGAGAGACTGGTGAAGAAGAGAAAGATGAAAATACAGAGAGAGCTAGAAGCAGCTGA
ATTCGTGAAGAAGCTCAAGTTGCTTCAGGATCAAGAACTCAAAAGAACTTGTAGATGT
AGAAAGAGAATTTATGGCCATGGCTGCAACAGAACACAATCCCGAGCCTGACATTTTGGT
GAACAATCAAAGCGGGAATACGAGATGTCAGCTTAACTCAGAGGACCTACTTGTGACGG
TGGCTCAATGGATGTAAATGGGAGGATAGAGATAGAGTAGAGCAAAACCGGTAACATAGC
AATAGAGAAGGTACCAAATCCCAAGGCTTGAGATCCGAAT

>G818 Amino Acid Sequence (domain in AA coordinates: 70-162)
MTAIPNVVDIESSSSLCQETATETVTVVERGSSDSSSKPDDVLLIKEEEDDAVNLSLGF
WKLHEIGLITPFLRKTFEIVDDKVTDPVWSWSPTRKSFIIWDSYEPSENLLPKYFKHKNF
SSFIRQLNSYGFKKVDSRWEFANEGFQGGKKHLLKNIKRRSKNTKCCNKEASTTTTETE
VESLKEEQSPMRLEMLKLKQQQEESSQHQMVTVQEKIHGVDTEQQHMLSFFAKLAKDQRFV
ERLVKKRKMKIQRELEAAEFVKLKLQDQETQKNLLDVEREFMAMAATEHNPEPDILVN
NQSGNTRCQLNSEDLLVDGGSMDVNGRIEIE*

>G849 (218..2077)

AACTCGAGAATTCTTCATTTCTTTTAAATCTTAGAATCTCGAGTTTTTGTATAAATCGAT
TCTAATTTTCCTTTGTACATTGTTTTATATATACATAAAACACACAAATCGGGTATGGG
GGAATTTGGGTTTTTAAGATAGCGTGATCTGTAATAATAAGTGGTTCGCGATCGTGATCAA
GAACTGGTGGCTGATAGTGATATGCATATTTGAGAGATGGTGTTCAGAGAAAAGTTAGA
TTGCCCTTCCCGTGGGATTTGATTTTCCCAACATTTCCAGAGCTCCTCGTTCATGCAGGAG
GAAGGTTCTAAACAAGAGGATTGATCATGATGATGATAACACTCAGATCTGTGCAATTGA
CTTACTAGCTTTGGCTGGAAGATTCTACAGGAAAGCGAGAGTTCTCTGCGTCTTCTAA
TGCAATTTGAAGAAATTAAGCAAGAGAAAAGTAGAAAATTGCAAGACTATTAAATCTGAGTC
TTCTGACCAAGGAAACTCTGTGTCAAAGCCTACTTATGATATCTCTACTGAGAAGTGTGT
GGTGAACAGTTGTTTTTCATTTCCCGATAGTGACGCGCTTTTGGAGCGGACTCCGATGTC
TGATTACAAGAAGATTTCATGGTTTGATGGATGTAGGGTGTGAAAAACAAGAAATGTAAATAA
TGGGTTTCGAGCAAGGAGAAGCAACCGATCGCGTGGGTGATGGAGGCTTAGTCACTGATAC
TTGCAACTTAGAGGATGCAACTGCGTTAGGTCTGCAGTTTCCGAAATCAGTCTGTGTGGG
TGGTGATTTAAAAATCACCATCCACCTTGGATATGACCCCTAATGGTTCCATGCTAGACA
TGGGAACCATACTAACCTAGGTAGAAAAGATGATGATGAAAAATTTCTATAGTTACCATAA
ACTTAGCAATAAAATTAAGTCGTATAGGTCTCCAACAATTCGAAGAATAAGAAAGTCCAT
GTCGTCCAAATACTGGAAACAAGTTCCAAAAGATTTTGGATACAGTAGAGCTGATGTGGG
TGTGAAGACTCTTTATCGCAAAAGAAAATCATGTTATGGTTACAACGCATGGCAGCGTGA
GATCATTTATAAGAGAAGAAGATCACCTGAAGAAGCTCGGTGTAACCTTCTGATGGAGG
ACTCAGTAGTGGAAGTGTTCCTCAAGTTACCCAAGAAGGGAGATACAGTAAAGCTAAGCAT
TAAGTCCTTTAGGATTCCAGAGCTTTTTATTGAAGTTCCAGAACTGCAACAGTAGGATC
ACTAAAGAGGACTGTGATGGAGGCTGTCACTGTTTTACTCAGCGGAGGAATACGTGTTGG
GGTGTAAATGCATGGAAGAAGGTTAGAGATGAAAGGAAAATCTGTCCAGACTGGGAT
CTCATGTGATGAAAATCTAGACAACCTTGGGTTCACTTGGAGCTAGTCCAGCAAAGT
TCCCTACCTTTGTGTTCTGAAGATCCTGCTGTGCCAACCGACCCTACAAGTTTGTCTGA
ACGGTCTGCGGCTCTCCTATGCTAGATTCTGGAATTCACATGCAGATGACGTGATTGA
TTCAAGAAATATTGTGGACAGTAACCTCGAATTAGTTCATATCAGGGTGACATATCTGT
TGATGAACCTTCATCAGATTCAAAGAGCTTGTCCCACTTCCAGAGTTGGAAGTCAAGGC
GCTTGCCATAGTTCCGTTGAACCAGAAACCTAAGCGTACTGAGCTAGCCAGAGGAGAAC
TAGGAGACCCCTCTCTGTGACAGAGGTAGAAGCTCTTGTAAGCAGTTGAGGAACTCGG
GACTGGAAGATGGCGTGATGTAAAATTGCGTGCTTTTCAGGATGCAGATCATCGGACTTA
CGTGGAATTGAAGGACAAATGGAAGACGCTAGTTCACACAGCAAGTATATCCCCACAGCA

ACGAAGAGGAGAGCCGGTGCCACAAGAACTGCTAGACAGAGTCTTGAGGGGCATACGGGTA
TTGGTCGCAGCACCAAGGAAAACATCAGGCGAGAGGAGCGTCCAAAGATCCAGACATGAA
CAGAGGTGGAGCTTTTGAATCAGGTGTTTCAGTGTAAAAAAGGAGGTACGCATTGGTGGG
TGGGTGTACAGAAGCAAACAACAATAAATGGACAACCTCAATTCTGCAAAGTTTAATT
GTCTTTATTTCTCGTTTTTTTTTTTTTCTCCTACATACACTTTTTTTTTTCT
>G849 Amino Acid Sequence (domain in AA coordinates: 324-413, 504-583)
MVFKRKLDCLSVGDFPNIPRAPRSCRKVLNKRIDHDDNTQICAIDLLALAGKILQES
ESSSASSNAFEEIKQKVENCKTIKSESSDQNSVSKPTYDISTEKCVVNSCFSFPDSGD
VLERTPMSDYKKIHGLMDVGCENKNVNGFEQGEATDRVGDGGLVTDTCNLEDATALGLQ
FPKSVCVGGDLKSPSTLDMTPNGSYARHGHNHTNLGRKDDDEKFYSYHKLSNKFYSYRSP
IRRIKSMSSKYWKQVPKDFGYSRADVGKTLRYKRKSCYGYNAWQREIIYKRRRSPDRS
SVVTSDDGLSSGSVSKLPKKGDTVLSIKSFRIPELFIEVPETATVGS�KRTVMEAVSVL
LSGGIRVGLMHGKKVRDERKTLSTQGISCDENLDNLGFTLEPSPSKVPLPLCSEDPVP
TDPTSLERSAASPMDSGIPHADDVIDSRNIVDSNLELVPYQGDISVDEPSSDSKELVP
LPELEVKALAIIVPLNQPKRTELAQRRTRRPFSVTEVEALVQAVEBELGTGRWRDVKLRAF
EDADHRTYVDLKDQKWLTVHTASISPPQRRGEPVPQELLDRVLRAYGYWSQHKGKHQARG
ASKDPDMNRGGAFESGVSV*
>G892 (21..1004)
TATAACAATTCTTCCAACAATGTCAATTGAGTCAGCCAATAACACGGACCGATAGTGCAC
CCAATGGAGCATTAGGACTTTTGGTCTCTACTGGTGCTACCATTTGTGATCGTATGGTCA
GAATTGCATCCTCTAACCCTATCAGAGATCGCCTGTCTCGATGTTTGAGGCAATTTGTCG
TTGAGATTGAAACGAGACAACGGCCTCGGTTTACTTTCAACCATGCTACTCCGCCTTTTG
ATGCTTCTCTGAGGCTCGTCTTCTCGAAGCTCTCTCGCTCATGTTTGAGCCTGCAACCA
TAGGTAGGTTTGGTGACAGACCCATTTCTTAGGGCAAGATCCAGAAACATCTTGGAACCTG
AATCAAGACCCCGACCGCAACATCGAAGACGACACAGCCTTGACAATGTTAACAATGGTG
GTTTACCTCTACCAAGAAGACATATGTTATTCTCCGGCCCAATAATCCGACTAGTCCAC
TCGGAAACATAAATTGCGCCACCAATCAAGCACCACCACGGCATGTGAACTCACATGATT
ACTTTACTGGAGCATCAAGCTTAGAGCAGCTGATTGAACAACATAACACAAGACGATAGGC
CTGGACCACCACCTGCGTCAGAACCACCATTAATTCCTACCATCTGTGAAAATAACAC
CACAACATCTAATAACGACATGTCCCAATGCACAGTGTGCATGGAAGAATTCATTGTTG
GTGGGGACGCAACGGAATTACCATGTAAACATATTTACCATAAAGATTGTATAGTCCCGT
GGCTTAGGCTTAACAATTCTTGCCCTATCTGCCGCGTGACCTGCCACTTGTCAACACCG
TTGCTGAATCTCGAGAAAGGAGCAATCCTATTAGACAAGACATGCCTGAAAGAAGGCGTC
CAAGGTGGATGCAACTCGGTAACATTTGGCCATTTAGAGCAAGATACCAAAGGGTTAGTC
CAGAAGAAACAGCAAACAGAAATCCTCGAGATAACAGGAGCTAACTCTGAATATTCCATG
GGAAATAAAATCGTGACTATCTATATGTATAGACTCTATGAGACATTGTCTATTTGAAT
GTGCATGTATATCTCAGAAATAAACTCAAGCGAAACATATTTAACGACTAAAAA
>G892 Amino Acid Sequence (domain in AA coordinates: 177-270)
MSLSQPITRTDSAPNGAFRTFGLYWYHCDRMVRIASSNPSEIACPRCLRQFVVEIETRO
RPRFTFNHATPPFDASPEARLLEALSLMFEPATIGRFGADPFLRARSNIPEPESRPRPQ
HRRRSLDNVNNGGGLPLPRRTYVILRPNNPTSPLGNI IAPPNQAPPRHVNSHDYFTGASS
LEQLIEQLTQDDRPGPPPPASEPTINSLPSVKITPQHLTNDMSQCTVCMEEFIVGGDATEL
PCKHIYHKDCIVPWLRLNNSCPICRDLPLVNTVAESRERSNPIRQDMPERRRPRWMQLG
NIWPFRRARYQVSPREETANQNPRDNRS*
>G961 (1..1200)
ATGTCAAAATCTATGAGCATATCAGTGAACGGACAATCTCAAGTGCCTCCTGGGTTTAGG
TTTCATCCGACCGGAGAGAGCTGTTGCAGTATTATCTCCGGAAGAAAGTTAATAGCATC
GAGATCGATCTTGATGTCTATTCGCGACGTTGATCTCAACAAGCTCGAGCCTTGGGACATT
CAAGAGATGTGTAAATAGGAACAACGCCACAAAACGACTGGTATTTCTTTAGCCACAAG
GACAAAAAATATCCGACGGGAACGAGAATAACAGAGCCACTGCCGCTGGATTTTGAAA
GCAACTGGCCGCGACAAGATCATATATAGCAATGGCCGTAGAAATTGGGATGAGAAAGACT
CTTGTTTTCTACAAAGGCCGAGCTCCTCACGGCCAAAAATCTGATTGGATCATGCATGAA
TATAGACTCGATGACAACATTATTTCCCCGAGGATGTACCGTTCATGAGGTGCTGAGT
ATTATAGGGGAAGCATCACAAAGACGAAGGATGGGTGGTGTGCTGATTTTCAAGAAGAAG
AATCTTCACAAAACCTTAAACAGTCCCGTCGGAGGAGCTTCCCTGAGCGGCGGCGGAGAT
ACGCCGAAGACGACATCATCTCAGATCTTCAACGAGGATACTCTCGACCAATTTCTTGAA
CTTATGGGGAGATCTTGTAAGAAGAGCTAAATCTTGACCTTTTCATGAACTCCCAAAC

CTCGAAAGCCCTAACAGTCAGGCAATCAACAACCTGCCACGTAAGCTCTCCCGACACTAAT
CATAATATCCACGTCAGCAACGTGGTCGACACTAGCTTTGTTACTAGCTGGGCGGCTTTA
GACCGCCTCGTGGCCTCGCAGCTTAACGGACCCACATCATATTCAATTACAGCCGTCAT
GAGAGCCACGTGGGCCATGATCATCTCGCTTGCCTTCCGTCCGATCTCCGTACCCACAGC
CTAAACCCGGTCCGCTTCGTACCACGCCGGTTTAAACAGGAATATACACCGGAGATGGAG
CTATGGAATACGACGACGTCGTCTCTATCGTCATCGCCTGGCCCATTTTGTACAGTGTCTG
AATGTTTTGTGCTTGTGTTGTCTCCTTCGTCTGCAGCTTCAGTTCTGGCCGTTCCAACCA
TGGCAGAGGCAGGTCATTTTCGATCTTTCATCGCCTCAGATGCAGATCTCTCTCCATTGA
>G961 Amino Acid Sequence (conserved domain in AA coordinates: 15-140)
MSKSMSISVNGSQVPPGFRFHPTEEELLQYYLRKKVNSIEIDLVDVIRDVDLNKLEPWDI
QEMCKIGTTPQNDWYFFSHKDKKYPTGTRTNRATAAGFWKATGRDKIIYSNGRRIGMRKT
LVFYKGRAPHGQKSDWIMHEYRLDDNIIISPEDVTVHEVVSIIIEASQDEGWVVCRIFFKK
NLHKTLSNPVGGASLSGGGDPKTTSSQIFNEDTLDQFLELMGRSCKEELNLDPFMKLPN
LESPNSQAINNCHVSSPDNENIHVSNVVDTSFVTSWAALDRLVASQLNGPYSYSITAVN
ESHVGHDLALPVSRSYPSPSLNRSASYHAGLTQXEYTPPELMELWNTTSSLSSSSPGPFCHVS
NVLLLVLCLRLQLQFQWPFQWQQRQVHFDLSSPQMQLSLH*
>G1465 (163..1125)
TATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTTCATTTGGAGAGGACACGC
TGACAAGCTGACTCTAGCTTATCTGGTACCGTCGACCTCATTCTTGCGTTTGATCTTTCT
TTCTCTAGATCCCATATTTTTCTTGATCAATTTAGTTTCATTATGGAGGAAGATGCAGCT
TTTGATCTACTCAAAGCCGAACCTCTAAACGCAGAAGACGATGCAATAATCTCACGTTAT
CTGAAGCGTATGCTCGTCAACGGAGACTCATGGCCTGATCACTTCATCGAAGACGCAGAC
GTGTTCAACAAGAATCCAAATGTGGAGTTCGATGCTGAGAGCCCTAGCTTCGTGATAGTT
AAACCTCGAACAGAGGCTTGTGGTAAAACCGATGGATGTGAAACTGGTTGCTGGAGGATC
ATGGGTGCTGATAAAACCGATAAAATCGACGGAGACTGTGAAGATTCAAGGGTTCAAGAAG
ATTCTCAAGTTCTGCCTAAAGAGGAAACCTAGAGGATACAAGAGAAGTTGGGTAATGGAA
GAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAGATCATGTGATTGCAAGATTCCGG
TTTATGTTTGAAGCTGAAATCAGTTTCTTGCTAGCCAAGCATTTCTACACTACATCAGAA
TCACTTCCTCGAAATGAGCTGTTGCCAGCTTACGGATTCTTTTCATCAGATAAGCAATTG
GAGGATGTATCTTATCCGGTGACGATAATGACTTCTGAAGGAAACGATTGGCCTAGCTAC
GTTACCAACAATGTGTATTGTCTGCATCCATTGGAGCTCGTTGATCTTCAAGATCGGATG
TTAATGATTACGGAACCTGCATCTTCGCTAACAAGACTTGTGGTAAAACCGATAGATGC
ATTAATGGTGGTTACTGGAATAATTTTGACCGTGATAGGCTGATCAAGTCAAAGTCCGGG
ATAGTTATTGGTTTCAAGAAGGTGTTTAAAGTTTCATGAAACGGAGAAAGAAAGATACTTC
TGTGGTGGAGAAGATGTGAAGGTAACCTGGACTCTAGAAGAGTATAGGCTTAGCGTGAAG
CAGAATAAATCTTGTGCGTTATCAAGTTTACTTATGATAACTAAGAATCTTTTCTTTGG
ATTTTATGATCATCTTAGTATCGCGACCGCTCTAGACAGGCCTCGTACCGGATCCTCTAG
CTAGAGCTTTCGTTCTGATCATCGGTTTCGACAACG
>G1465 Amino Acid Sequence (conserved domain in AA coordinates: 242-306)
MEEDAADFLLKAEELNAEDDAIISRYLKRNVVNGDSWPDHFIEDADVFKNPNVFEDEAS
PSFVIVKPRTEACGKTDGCETGCWRIMGRDKPIKSTETVKIQGFKILKFCLKRKPRGYK
RSWVMEYRLTNLNLWKQDHVICKIRFMFEAEISFLAKHFYTTSESLPRNELLPAYGFL
SSDKQLEDVSYPTIMTSEGNWPSYVTNNVYCLHPLELVDLQDRMFNDYGTCTIFANKTC
GKTDRCCINGGYWKILHRDLIKSKSGIVIGFKKVFKEHETEKERYFCGGEDVKVWTWLEE
YRLSVKQNKFLCVIKFTYDN*
>G425 (45..1196)
GAAACAGTCTTCTCTTCTCCGATCCCAAAAACGCAGGAAAACAATGTGCTTTAACAGCTCCC
ACCTCCTTCTCTCAACAAGAACCTTCTCTCCGACACTTCACCGATCAATCACAGCAACCTC
CGCCGACGCTCACTTCTCTGAAACACCTTCGCTTGTCAACGCCAGTTTCTCAACCTCCCTA
CCACCCTTACCAGTGGCGATTCCGATCTCGCTCCTCCGACCGCAACGGAGACAATTCCGTT
GCTGATACAAAACCCACGCTGGCTCTCCTTTCATTCCGAGATGCAAAATACTGGAGAAGTACG
TTCTGAAGTTATCGACGGAGTCAACGCCGATGGTGAACGATACTCGGCGTTGTAGGAGGT
GAAGATTGGCGGAGTGTAGCTATAAGCGCGCATTTTAAGACATCCGATGTACGAGCAGC
TTCTTGCGGCTCACGTGGCTTGCCTTAGGGTTGCGACTCCCGTTGACCAGATTCCGAGGATC
GATGCTCAGCTCAGTCAGTTGCATACCGTCGCCGCGAAATACTCCACTCTTGGTGTGGTTGTT
GACAACAAGGAACCTTGATCATTTTCATGTACATTATGTTGCTTGTATGTTTAAAGAAC
ACTCCAACACCACGTTTGTGTCCATGCAATGGAAGCCATTACGGCTTGTGGGAGATTGAACA

ATCACTGCAATCCCTAACTGGAGTTTCTCCAAGTGAAAGTAATGGTAAGACAATGTCGGATGA
TGAAGATGATAATCAAGTAGAGAGCGAGGTGAACATGTTTGATGGAAGTTGGACGGTTCAG
ATTGCTTGATGGGGTTTGGTCCTCTTGTTCACCGAGAGAGAGATCTTTGATGGAACGTG
TGAAGAAAGAACTGAAGCATGAGCTTAAACAGGGTTTCAAAGAGAAGATTGTGGACATAAG
AGAAGAGATAATGAGGAAGAGAAGGGCGGGAAGTTGCCAGGAGATACGACTTCTGTACT
CAAAGAATGGTGGCGAAGTCACTCGAAATGGCCATACCCAAGTGAAGATAAGGCAAAA
GAACTGGAACAGCAACTCTTCCACGTCATCTACTCTACCAAGAACAACGTAAACGGACC
GGGAAGTCGTAGGTGACATAGCGGCTAACTAGAGGATGGTTCTTTGCCATGTGAATTCTTGG
GAACCGTATATGAAAGAAACGAATCCGGTTCTATGCTCGTACAGAGTGTGTTATTTGTATAGT
GGATACCGGTTAGCCTATGAAACCGGATTCGGAGTCCAAATTGTTGTTTGTAAACGACTTAGT
AGTTTTTGAAGTGATCTGTTTCGTTGGTTTGGCTCTTGTAAACGACGCTTAAGCAAGTGTGGG
TTTTTCTTGTAAAGTGTCAATATGTTTCGTTTGTAAATGAATGTATCAAGCAATATTTATCATAATT
AACTAGCTTGAAATGTAAAAA

>G425 Amino Acid Sequence (domain in AA coordinates: TBD)
MSFNSSHLPPQEDLPLRHFTDQSQQPPQRRHFSETPSLVTASFLNLPPTLTADSLLAPPHR
NGDNSVADTNPRWLSFHEMQNTGEVRSEVIDGVNADGETILGVVGGEDWRSASYKAAILR
HPMYEQLLAAHVACLRVATPVDQIPRIDAQLSQLHTVAKYSTLGVVVDNKLDFHMSHYVVL
LCSFKEQLQHHVCVHAMEAITACWEIEQSLQSLTGVSPSESNKMTSDDEDDNQVESEVNM
FDGSLDGSDDLGMFGPLVPTERERSLMERVKKELKHELKQGFKEKIVDIREEIMRKRRAGKLP
GDTTSLKIEWRTHSKWPTYPTEDKAKLVQETGLQLKQINNWFINQRKRNWNSNSSTSSTLT
KNKRKRTGKS*

>G347 (1..570)

atgaaagtagcagatatgcaggaccagctgggtgtgcatgggtgtaggaatttattgatg
taccctagaggagcatctaattgtgcgttggtgcgttatgtaacactatcaacatgggtcct
cctcctcctccacctcacgacatggcacacattatatgtggtgggtgtagaacaatgctt
atgtatacgcggtggggttagtagcgttaagatgctcctgctgtcaaaactacgaaccttggtg
ccagcgcactccaatcagggttgcccatgctcctccagtcaggttgccgcagatcaattgt
gggcattgtcgagcagacctcatgtatccttacggtgcatcatccgtcaaatgcgctggt
tgtcaattcgtaactaacgttaatatgagcaatggaagggtacctctcccaactaaccgg
ccaatggaacagcttgccccctctacatcaacttcaacaccacctctcagacccaa
accggtgtgtgtagaaaaccccatgtccggttgatgaaagcggaaggttggtgagcaatggt
ggtgtgagtgagcaactgacaaaaagtaa

>G347 Amino Acid Sequence (domain in AA coordinates: 9-39, 50-70, 80-127)

MKVADMQDQLVCHGCRNLLMYPRGASNVRCALCNTINMVP PPPPPHMAHIICGGCRTML
MYTRGASSVRCSCCQTTLNLPVPAHSNQVAHAPSSQVAQINCCHCRTTLMYPYGASSVKCAV
CQFVTNVNMSNGRVPLPTNRPNGTACPPSTSTSTPPSQTVVVENPMSVDESGLVSNV
VVGVTDDKK*

>G1512 (1..732)

ATGGAAGGGAACCTTCTTCATCAGGTCTGATGCTCAACGAGCATGACAATGGCTTCATA
GCCAAACAAAAACCTAATCTCACCACGGCTCCAACAGCAGGTCAAGCTAATGAAAGTGGC
TGTTTGGACTGCAACATCTGTTTAGACACAGCCCATGATCCGGTGGTCACTCTCTGCGGG
CACCTTTTCTGCTGGCCTTGCAATTTACAAGTGGTTACATGTTCAATTATCTTCTGTCTCC
GTTGATCAGCACCAGAACAAATGCCCTGTTTGTAAATCCAACATTACTATCACCTCTTTG
GTTCTCTCTATGGAAGAGGCATGTCTTCGCCTTCTTCCACGTTTGGCTCCAAGAAACAA
GACGCACTGTCCACTGACATACCCCGCAGACCTGTCTCCATCAGCCTTACGCAATCCGATT
ACCTCAGCATCATCTCTGAACCAAGCTTGCAACATCAAACCTCTGTCTCCTTCATTTTCA
AATCATCAGTATTTCCCTCTGCTGGCTTCAACACAACCGAATCAACCGACCTTGCCAATGCT
GTAATGATGAGTTTCTCTACCTGTGATTGGAATGTTTGGAGACCTGGTCTACACCAGG
ATATTCGGGACCTTCACAAACACAATAGCTCAGCCTTACCAAAGCCAGAGGATGATGCAG
CGTGAGAAGTCTCTTAATCGGGTATCGATATTCTCTCTTGTGTCATCATCTTTGCCTC
CTTCTCTTCTAG

>G1512 Amino Acid Sequence (domain in AA coordinates: 39-93)
MEGNFFIRSDAQRADHNGFIAKQKPNLTAPTAGQANESGCFDCNICLDTAHPVVTLCG
HLFCWPCTIYKWLHVQLSSVVDQHQNCPVCKSNITITSLVPLYGRGMSSPSTFGSKKQ
DALSTDIPRRPAPSALRNPIITASLNLPSLQHQTLSPSFHNHQSYPGFTTTTESTDLANA
VMMSFLYPVIGMFGDLVYTRIFGTFTNTIAQPYQSQRMMQREKSLNRVSIFLCCIILCL
LLF*

>G2069 (1..1026)

ATGGAAGGAGGAGGAAGAGGACCAATCAAACGATTCTCAGTGAAATAGAACATATGCCT
GAAGCTCCACGTCAACGTATCTCTCATACCGTCGAGCTCGCTCTGAAACCTTCTTCTCC
GGCGAATCAATCGACGATCTCCTCTTATTCGATCCTTCCGATATCGATTCTCTTCTCTA
GACTTCTCTCAACGCTCCACCACCACCACAACAATCACAACAACAACCGCAAGCTTCTCCC
ATGTCCGTTGATTTCGGAAGAAACCTCATCGAACGGTGTGTTCTCTCCTAATTCTCTTCTT
CCAAAACCCGAAGCTAGATTTCGGTCGCCATGTTTCGTAGCTTCTCGGTTGATTCCGATTTC
TTCGATGATTGGGTGTTACTGAGGAGAAAGTTTATAGCTACAAGTTCAGGAGAGAAGAAG
AAAGGGAATCATCATATAGCAGGAGTAATTCTATGGATGGAGAGATGAGTTCGGCGTCG
TTTAATATCGAATCGATTCTTAGCTTCTGTGAGTGGTAAAGATAGTGGGAAGAAGATATG
GGTATGGGTGGTGATAGACTTGCTGAGCTTGCTTTGCTTGATCCTAAAAGAGCTAAAAGG
ATTTTAGCGAATAGACAATCTGCGGCGAGGTCGAAAGAGAGGAAGATTAGGTATACTGGT
GAGTTAGAGAGGAAGGTTAGACACTTCAGAAATGAAGCTACTACATTGTCTGCTCAAGTC
ACTATGTTACAGAGAGGAACATCAGAGCTGAACACTGAAAATAAACACCTCAAAATGCGG
CTTCAAGCTTTAGAGCAACAAGCTGAACCTTAGGGATGCTTTGAATGAAGCGCTGCGGGAT
GAAGTGAACCGACTTAAGGTGGTAGCTGGAGAAATTCCTCAGGGGAATGGAATTCCTTAC
AACCGTGTCTCAATTCTCATCTCAGCAATCGGCAATGAATCAGTTTGGGAACAAAACGAAC
CAACAGATGAGTACAAACGGGCAGCCATCGCTCCCAAGCTACATGGATTTACCAAGAGA
GGCTGA

>G2069 Amino Acid Sequence (domain in AA coordinates: TBD)
MEGGGRGNPTILSEIHMPEAPRQRISHHRRARSETFFSGESIDDLFFDPSDIDFSSL
DFLNAPPPPPQSQQQPQASPMVDSEETSSNGVVPNSLPKPEARFGRHVSFSVDSDF
FDDLGVTEBKFIATSSGEKKKGNHHHSRNSMDGEMSSASFNIESILASVSGKDSGKKNM
GMGGDRLABLALLDPKRAKRILANRQSAARSKERKIRYTGELERKVQTLQNEATTLSAQV
TMLQRGTSELNTENKHLKMRLLQALEQQALRDALNEALRDELNRLKVVAGEIPQGNNGSY
NRAQFSSQQSAMNQFGNKTNQMSSTNGQPSLPSYMDFTKRG*

>G1852 (55..1857)

CATCTGATCTGCTCTCGAAGACGAAAGCTTCGAGTACTGGTTGAAGCTAAAGCTATGGGA
CACGTGAATCTACCTGCATCAAAGCGTGGTAACCCCTCGTCAATGGCGTCTCCTCGACATC
GTAACCGCTGCTTTCTTCGGTATCGTACTTCTCTTCTTCATCCTTTTATTCATCCTCTT
GGTGATTCCATGGCGGCTTCTGGTCGGCAAACGCTGCTTCTCTTACGGCGTCAGATCCG
AGGCAACGGCAGCGATTAGTGACTTTGGTTGAAGCTGGTCAGCATTGCAACCGATCGAG
TATTGTCCTGCGGAAGCTGTTGCTCATATGCCTTGTGAGGATCCGAGAAGGAATAGTCAG
CTTAGTAGAGAGATGAATTTCTATAGGGAGAGACATTGCTCTTGCCTGAGGAGACTCCG
CTCTGTTTGATTCTCTCCGCTTCTGGTTATAAAATTCCTGTTCCGTGGCCTGAGAGTCTT
CACAAGATTTGGCATGCAACATGCCATATAACAAAATGCTGACCGGAAAGGTCATCAA
GGATGGATGAAAAGGGAAGGGAATACTTTACTTTCCAGGCGGTGGCAGCATGTTTCTCT
GGCGGAGCTGGCCAATACATTGAAAAGCTTGACAGTATATTCCGCTTAATGGTGGAAC
TTGAGAACTGCTCTTGACATGGGATGCGGGGTAGCTAGTTTGGAGGTAATCTACTATCT
CAAGGCATTCTAGCCCTCTCATTTGCTCCAAGAGATTACATAAATCTCAAATTCAGTTC
GCTTTGGAAAGAGGAGTGCTGCATTTGTTGCCATGCTTGGCACTCGTAGACTCCCTTT
CCTGCATACTCCTTTGACCTGATGCACTGTTCCCGATGTTTGGATCCTTTTACGGCTTAC
AATGCAACTTACTTTCATCGAAGTAGATAGGTTACTGCGCCCTGGAGGATATCTTGTATC
TCTGGCCACCTGTACAAATGGCCTAAACAAGACAAAGAATGGGCTGATCTTCAGGCGGTG
GCTAGAGCTTTGTGCTATGAGCTAATTGCGGTTGATGGAAACACTGTCATCTGGAAGAAG
CCTGTTGGAGATTATGCTTACCTAGCCAGAATGAGTTTGGGCTTGAGTTGTGTGATGAG
TCTGTTCCGCCAAGTGATGCATGGTATTTTAAATTGAAGAGGTGTTTACCAGGCCATCA
TCCGTCAAAGGAGAAACCGCTTTGGGAACATATCCAAGTGGCCGGAGAGGCTTACTAAA
GTTCTCTTCTAGGGCCATTGTGATGAAAAACGGATTGGATGTGTTTGAAGCAGATGCAAGG
CGGTGGGCAAGACGCGTTGCTTATTACAGGGATTCTCTTAACCTGAAGCTGAAATCTCCA
ACTGTCCCAATGTGATGACATGAACGCATTCTTCGGAGGCTTTGCAGCAACCCCTTGCA
TCTGATCCTGTGTGGTTATGAATGTCATTCCAGCTCGGAAGCCATTAATCTTGACGTG
ATTTATGACAGAGGTTCTCATCGGTGTTTACCATGATTGGTGTGAACATTTTCAACATAT
CCCCGCACGTATGATTTTCATCCATGTATCAGGAATTGAATCACTGATAAAACGACAAGAC
TCAAGCAAATCGAGGTGTAGCCTAGTAGATCTAATGGTAGAGATGGACAGAATATTACGT
CCAGAAGGAAAGGTTGTGATCCGAGACTCTCTGAGGTGCTAGATAAAGTCGCACGAATG
GCTCATGCTGTAAGATGGTCTTCTTCCATACACGAGAAAGAACCTGAATCCCATGGAAGA

GAGAAGATTCTTATCGCAACCAATCTCTCTGGAAATTGCCATCAAACCTCCCACTGAAGA
CACAAAAGAAGAAGAAAAGAAGCTCTTCTCAATCTTGTAGGTACTGTCACTTGCTCT
CCAGCCC.

>G1852 Amino Acid Sequence (domain in AA coordinates: 1-601)
MGHVNLPAKRGNNPRQWRLLDIVTAFFGIVLLFFILLFTPLGDSMAASGRQTLTLLSTAS
DPRQRQLVLTLEAGQHLQPIEYCPAEVAHMPCEDPRRNSQLSREMNFYRERHCPLPEE
TPLCLIPPPSGYKIPVPWPESLHKIWHANMPYNKIADRGHQGWMMKREGEYFTFPGGGTM
FPGGAGQYIEKLAQYIPLNGGTLRTALDMGCGVASFGGTLTLLSQGILALSFAPRDSHKSQI
QFALERGVPAFVAMLGTRRLPFPAYSFDLMHCSRCLIPFTAYNATYFIEVDRLLRPGGYL
VISGPPVQWPKQDKEWADLQAVARALCYELIAVDGNTVIWKKPVGDSCLPSQNEFGLELC
DESVPPSDAWYFKLRKCVTRPSSVKGEHALGTISKWPERLTKVPSRAIVMKNGLDVFEAD
ARRWARRVAYYRDSNLNKLKSPTVRNVMDMNAFFGGFAATLASDPVWVMNVI PARKPLTL
DVIYDRGLIGVYHWDCEPFSTYPRTYDFIHVSGIESLIKQDSSKSRCSLVDLMVEMDRI
LRPEGKVVIRDSPEVLDKVARMAHAVRWSSSIHEKEPESHGREKILIIATKSLWKLPNSH
*

>G1793 (59..1783)
AGTGATTTATTGATTACCCCAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
GAATTTCTAACAACTGGCTTGGCTTTCTCTTTTACCGAACAACTCTTCTTTGCCTCTCA
TGAATACAACCTTGGCTTGGTCAGCGACCATATGGACAACCC'TTTTCAAACACAAGAGTG
GAATATGATCAATCCACACGGTGGAGGAGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
CGATTTTCTCGGTGTGAGCAAACCGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
CGACTCAGACTACTACTTCCATACCAATAGCTTGTATGCCTAGCGTCCAATCAAACGATGT
CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACAAACAGTAGCTATCATGAGCTTCAAGA
GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
TGTAGACAAAGCTTACCATCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAAGACATCGATGGACTGGTCGATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACTTAAGTACTGGGGTCTCTTC
AACTACTACTAATTTCCCCATTACAACTACGAGAAAAGAAGTAGAGGAAATGAAGCACAT
TACGAGACAAAGACTTCGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
CCGAGTCGCCGGAACAAAGACCTCTACTTGGGAAC'TTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCCATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCCGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCAGCTTCAACCTTACCCTCTAAGCATTCAACAACCATTAGAGCC
TTTCTATCTCTTTCAGAACATGACATCTCTCATTACAACAACAACATGCTCACGATTC
CTCCTCTTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTGCAGCAACAGTCGAGCCAGAACTCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCACTGGACTTGTCTCTACCTCTATCGTTGACAACAATAATA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACCACGGTATTGG
TATTGGGTCCAGCTCGACTGTGGATCGACCGAGGAGTTTCCAACCGTTAAAACAGATTA
CGATATGCCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTTGGACCACTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGT'TTTCATATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)
MNSNNWLGFPLSPNNSSLPPEYNLGLVSDHMDNPFQTQEWNMINPHGGGGDEGGEVPKV
ADFLGVSKPDENQSNHLVAYNDSYFHTNSLMPSVQSNVVAACDSNTPNNSSYHELQ
ESAHNQLSLTSLMGTTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTGRYEAHLWDNSCRREGQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGP
STTNFPITNVEKEVEEMKHMTRQEFVAAIRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFFSTEEEAEEAYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESSRKREAEMIALGSSFYGGGSSSTGSGSTSSRLQLQPYPLSIQQPLE
PFLSLQNNDISHYNNNNNAHDSSSFNHHSYIQTQLHLHQQTNNYLQQQSSQNSQQLYNAYL
HSPALLHGLVSTSIVDNNNNNGGSSGSYNAAFLGNHIGIGSSSTVGSSTEEFPTVKTD

YDMPSSDGTGGYSGWTSSESQVGSNPGGVFTMWNE*
>G761 (521..1549)
GGGGCCGACCGCCGCCCGGGCAGGTCTAGGTTCAAAAGGACTCACAAGAGAGAGATAGT
ATGATTGATAGGGAAAGAGAGAGAGATGAAAGAAAGTAAATATATAATAGATTATTAGG
ACACGAGTGTCTATCTTTTGATTTGTGTCTTGTGTGCTCTCTTTCTTCTTCTCCTCGAA
TGATCATCTTTATATAACCCCTACTCTCTTTCTCTTTTCCCATTTCTTTCATATCATTCTCC
CTTCTCTCTCGGGATCTGATCTCTCTTTCCAGTAACCTATTCCCGAGGAGCACTGTCAA
ATCTTGTCCACTCTTTGATCTTATCTCGATCTCTTTCTCTTTCTAGTCTTGTGTAGTCTT
CAAACCTTGTGATGTTATCTATATAGTAATCACGAGAGAGAATCATACAATAGCTGAAACA
TAAAGCTTTCTTAGAAGCTTTAAAAAGGTCTCATCTGGATTATCCTGTTAATTTCTAGA
GTTTCTTCAGGCAGATTATTAACCGATCAAGAAGACAAACATGAATTCATTTTCCACGT
CCCTCCGGGTTTATAGATTTACCCGACAGATGAAGAACCTGTAGACTACTACCTGAGGAA
AAAAGTCGCATCGAAGAGAATAGAAATTGATTTCAATAAGGACATTGATCTTTACAAGAT
TGAGCCATGGGACCTTCAAGAGTTGTGCAAAATTGGGCATGAAGAGCAGAGTGATTGGTA
CTTCTTTAGCCATAAAGACAAGAATATCCACAGGACTCGAACCAATAGAGCAACAAA
AGCAGGGTTTGGAAAGCCACCGGAAGAGATAAGGCTATCTATTTGAGGCATAGTCTAAT
TGGCATGAGGAAAACACTTGTGTTTTACAAGGGAAGAGCCCCAAATGGACAAAAGTCTGA
TTGGATCATGCAGAAATACCGCTTAGAAACCGATGAAAACGGAACCTCCTCAGGAAGAAGG
ATGGGTTGTGTGTAGGGTTTCAAGAAGAGATTGGCTGCAGTTAGACGAATGGGAGATTA
CGACTCATCCCCTTACATTGGTACGATGATCAACTTTCTTTTATGGCTCCGAGCTCGA
GACAAACGGTCAACGACGGATTCTCCCAATCATCATCAGCAGCAGCAGCAGCAGCAGCA
ACAACATATAGCCATATGGCTCAATGCATCTGCTTACGCTCTCAACAACCTAAGTTGCA
ATGCAAGCAAGAGCTAGAACTACACTACAACCACCTGCAATCAAAATATCGCGCATGAGGA
ACAATTGAATCAAGGAAATCAGAACTTCAGCTCTCTATACATGAACAGCGGCAACGAGCA
AGTGATGGACCAAGTCACAGACTGGAGAGTTCTCGATAAAATTTGTTGCTTCTCAGCTAAG
CAACGAGGAGGCTGCCACAGCTTCTGCATCTATACAGAATAATGCCAAGGACACAAGCAA
TGCTGAGTACCAAGTTGATGAAGAAAAAGATCCGAAAGGGCTTCAGACATGGGAGAAGA
ATATACTGCTTCTACTTCTTCGAGTTGTCAGATTGATCTATGGAAGTGAGCTGAAAGAGA
AGACATATAAATGCATATATACATATATATACGTACACAGCAACTAATCAAGTG
TAGATGATGATGATGGTACAGATTATATATTGCTTTGATTGATTCTTACTACATTATTGA
ACTTATGTCATATGCATATATACATTGCGTATCTATGCATATTTATACTTGTACTCAATA
TGATTAACCATATATAAATCTAATCTAAATGTAACCTCAATATTTTAAATAGACAAT
TGTCTCTTCTTATTAGAAAAAAA
>G761 Amino Acid Sequence (domain in AA coordinates: 10-156)
MNSFSHVPPGFRFHPDDEELVDYLLRKKVASKRIEIDFIKDIDLYKIEPWDLQELCKIGH
EEQSDWYFFSHKDKKYPGTGRTNRATKAGFWKATGRDKAIYLRHSLIGMRKTLVFKGRA
PNGQKSDWIMHEYRLTDENGTPOEEGWVVCVFKRLAAVRRMGDYDSSPSHWYDDQLS
FMASELETNGQRRILPNHHQQQHEHQHMPYGLNASAYALNPNLQCKQELHLHYNHLQ
SNIAHEQLNQGNQNFSSLYMNSGNEQVMDQVTDWRVLDKFVASQLSNEEAATASASIQN
NAKDTNSAEYQVDEEKDPKRASDMGEEYTASTSSSCQIDLWK*
>G1056 (10..798)
GCTACATATATGGGTTCTATTAGAGGAAACATTGAAGAGCCTATATCTCAGTCATTAACG
AGGCAGAACTCTCTCTATAGCTTAAAGCTCCATGAGGTTCAAACCCACTTAGGAAGTTCT
GGAAACCACTAGGAAGCATGAACCTTGATGAGCTTCTCAAGACTGTCTTGCCACCAGCT
GAGGAAGGGCTTGTTCGTGAGGGAAGCTTGACGTTACCTCGAGATCTCAGTAAAAAGACA
GTTGATGAGGTCTGAGAGATATCCAACAGGACAAGAATGGAAACGGTACTAGTACTACT
ACTACTCATAAGCAGCCTACACTCGGTGAAATAACACTTGAGGATTTGTTGTTGAGAGCT
GGTGTAGTGACTGAGACAGTAGTCCCTCAAGAAAATGTTGTTAACATAGCTTCAAATGGG
CAATGGGTTGAGTATCATCATCAGCCTCAACAACAAGGGTTTATGACATATCCGGTT
TGCGAGATGCAAGATATGGTGTGATGAGGTGGATTATCGGATACACCACAAGCGCCTGGG
AGGAAAAGAGTAGCTGGAGAGATTGTGGAGAAGACTGTTGAGAGGAGACAGAAGAGGATG
ATCAAGAACAGAGAATCTGCAGCAGCTTACGAGCTAGGAAACAGGCTTATACATGAA
TTAGAGATCAAGGTTTCAAGGTTAGAAGAAGAAACGAAAACTTCGGAGGCTAAAGGAG
GTGGAGAAGATCCTACCAAGTGAACCACCACCATCTAAGTGAAGCTCCGGCGAACA
AACTCTGCTTCTCTGATCTTAAAGACTCTTCTTTCTTCTTCTTCTTGTGTTGGTTT
ATATCAGACCGCTTGTCTTGTATATTGTGTAGACTTTATTGACTTTGAACAGCATGT
CTTTATAACATTTCTTGAGTGT

>G1056 Amino Acid Sequence (domain in AA coordinates 183-246)
MGSIRGNIEEPISQSLTRQNSLYSLKLHEVQTHLGSSGKPLGSMNLDELLKTVLPPAEEG
LVRQGSLLTPRDLSSKKTDEVWRDIQQDKNGNGTSTTTTHKQPTLGEITLEDLLLRAVV
TETVVPQENVNIIASNGQWVEYHHQPQQQQGFMTYPVCEMQDMVMMGGLSDTFPQAPGRKR
VAGEIVEKTVERRQKRMIKNRESAARSARKQAYTHELEIKVSRLEEENEKLRLKEVEK
ILPSEPPDPKWKLRRTNSASL*

>G1447 (82..1086)

AAAAACCCCTAACCTAATCTCTCAAGACAACCTCAAAGGTCTCTCCTTTTATAGGTTTAT
TATCACTTCCGTATAATCGCCATGTCTTCTTACCATGGAAAAAACAAAATCGAGTCGA
ATCTTAAGATTCATTTCTGAGTTTCAACAATCACCGTTCGTTGAAACTGGCTTTCCAAC
TCTCTGATCGATCTCTTCTCAAGAATCGCGATCGTCTAAAAAATCTCCATCTAAACGC
TTCCAACGAATCGAACGCCAGATTGAAACCGCTCCAAACGCTTCTCGTTGAGTAATCAA
GATACGATTTTGGAAAGCCCTCGAGGATTAAAAACGTTGAAAGTAAGTTCGAGAAAGTT
AATTGCGTTAAAGGTAAATCAGCGCGCTTGAAGAAGAACGCGATTAAAAATAGCGTTTTC
GGCGGTAGCCGTGAGGTCGTTTGGATGGCGTTTAAAGTTTGGATAGTAGCGTTGCTCGCC
TTGAGCACGAAGAAGAAGCTCACTTTAGGAATCACTCTCTGCGCTTCGCTCTTCTCTTA
ACAGAGCTCGTGGCGCGCGTGTTCACGCGCTCTAATAACACCGACAAAGACAAAAAC
GCGATTGCCCGCGAGAAAAATCGAACTTTTGGATGAAACTCGAGTTCCCAAGCGATTCCA
TGTCTGAGGAAACAGAGCATGTAGTATCTGAAACAGAGGTTTCGAAAGTGAAGGTTTA
ACGATACGTGATCTGTTGTCAAAGGACGAGAAATCAACAAGTAAAGTTGGAGACTAAA
TCGAAGATTGTGAAGAAGTTGAGGAGTTACAATAAGAAGGATAAGAAGACGATGAAGATC
AAAGAAGAGCTTTTGGATTGAAGTCTCGAGTTTGGTTTTAGAAGATAAACCAAAGAAAT
GAGTCTGAGAGACGAAGAAGAAACGTTGAATCCTCCAGTGGTTGGATCAAACCTGAAT
GGGATTGTTCTGATCGTGATGTGCTAACCGGTTTGTATGTGGGAAGGTCTTAGCTATT
GTTCTGACACTATCATGTTTGGTTCTTAGATTAGGAGCAGTCAAAAAAGTTAATCTTTC
ATATAATTTTTTTTGTATTTTTTAACATGCTTGCATGTGAACTGTAAATTTTTCTCATT
CATATGAAGGAGATTGGATTGAATGTTGAATACTAA

>G1447 Amino Acid Sequence (domain in AA coordinates: 3-54, 124-156)

MSSLPWKKPKSSRILRFISEFQQSPFVETGFPSTLIDLFFKNRDLKKSPSKRFQRIERQ
IRTAPNASSLSNQDTIFEKPSRIKTVRSKVEKVNVCVKGSAAALKNAIKNSVFGGSGEVV
LMAFKVLIVALLALSTKKKLTGLITLSAFALLLTELVAARVFTSRNNTDKDKNAIAREKI
ETFDETRVPAIPCEBTEHVSETEVSKLKLGLTIRDLSSKDEKSTSKSWRLSKIVKKL
RSYNKKDKKTKMKIEESLIEVSSLVLEDPKKIESERDEEETLNPVVGSNLNGIVLIVI
VLTGLLCGKVLAIVLTLSCLVLRGLGAVKKVNLCI*

>G323 (77..826)

CTGCTCATATCAGCCATTGACACAGTTGCTTTGGGTTTCCCTCAAACGGCGCCGATTGTC
TGGATTTTGACCACTGATGCGCTTAGATCAATCTTTTGAAGATGCTGCTTTACTTGGAGA
ACTCTATGGAGAAGGTGCATTTTGTTCAGAGCAAGAAACCTGAACCCATTACAGTCTC
GGTTCCTTCTGATGATACTGATGATTGCAATTTTGAAGTCAATATTTGCTTAGACTCGGT
GCAAGAACTGTTGTGACTCTCTGTGTCACCTCTTTTGTGCGCTTGTATTACAAATG
GCTTGATGTACAGAGCTTCTCAACAAGTGATGAATACCAAAGACATAGACAGTGTCTGT
TTGTAAATCTAAAGTTTCTCATTCTACTTTGGTTTCTTTGTATGGTAGAGGCCGTTGTAC
TACTCAGGAGGAAGGTAAAAACAGTGTGCCTAAAAGACCCGTAGGACCGGTTTATCGGCT
TGAAATGCCGAATTCACCTTATGCAAGTACTGATCTGCGGTTATCACAACGGGTTTATC
CAATAGCCACAGGAAGGTTACTACCTGTCTCAGGGGTGATGAGCTCGAACAGTTTATC
ATACTCTGCTGTTTTGGATCCGGTGATGGTGATGGTTGGAGAAATGGTAGCTACGAGGT
GTTTGGAAACACGAGTGATGGATAGATTGCGTATCCGGACACTTACAATCTCGCAGGGAC
TAGCGGGCCGAGGATGAGAAGGCGGATAATGCAGGCAGATAAATCGCTGGGAAGAACTTT
CTTCTTCTTTATGTGTTGTGTTGTTCTGTGTCTTCTTGTGTTTATAGGTTTTCATAGCTAG
CTTGGTTCTGCTACTGTTCAAGTTTCTTCAGG

>G323 Amino Acid Sequence (conserved domain in AA coordinates: 48-96)

MALDQSFEDAALLGELYGEGAFCKSKKPEPITVSVPSDDTDDSNFDCNICLDSVQEPVV
TLCGHLFCWPCIIHKWLDVQSFSSTSEYQRRHQCPVCKSKVSHSTLVPLYGRGRCTTQEEG
KNSVPKRPVGPVYRLMPNSPYASTDLRLSQRVHFNSPQEGYYPVSGVMSSNSLSYSAVL
DPVMVMVGEMVATRLFGTRVMDRFAYPDTYNLAGTSGPRMRRIMQADKSLGRIFFFMC
CVVLCLLLF*

>G176 (41..1606)

AGAAGAAGAAGAAGAGTACCTCATACGTAAACCATTGATGGGCTCTTTTGATCGCCA
AAGAGCTGTTCCGAAATTCAAAACAGCAACACCGTCACCGCTCCCTCTTTCTCCTTCGCC
TTACTTCACTATGCCTCCTGGCCTTACTCCCGCCGACTTTCTCGACTCTCCTCTTCTCTT
CACTTCCCTCCAAACATTTTGCCGTCTCCTACGACAGGCACATTCCAGCGCAATCTCTGAA
CTATAACAATAACGGTTTGTCTATTGACAAAAATGAAATCAAATATGAAGACACAACCTCC
TCCCTTGTTCCTACCATCTATGGTAACTCAGCCTTTACCTCAACTGGATTATTCAAATC
CGAAATCATGTGCGAGTAACAAAACCTCTGATGACGGCTACAATTGGCGCAAATACGGGCA
GAAGCAAGTCAAAGGAAGCGAAAACCCGAGGAGTTACTTCAAATGCACGTATCCAAATTG
TCTCACAAGAAGAAAGTAGAGACGTCTCTTGTAAGGGTCAGATGATTGAGATTGTCTA
TAAAGGAAGCCACAATCATCCCAAGCCCCAATCCACGAAGCGATCATCTCCACCGCTAT
AGCAGCACATCAGAACAGCAGTAATGGAGACGGTAAAGACATTGGTGAAGATGAAACAGA
GGCCAAGAGATGGAAGAAGAGAAGATAATGTGAAGGAGCCAAGAGTGGTGGTTTCAGACAAC
AAGTGATATAGACATTCTTGACGATGGCTACAGATGGAGAAAGTATGGTCAGAAAGTCGT
CAAGGGTAATCCAAATCCAAGGAGCTATTACAAGTGCACATTACAGGATGTTTTGTAAG
GAAACACGTTGAAAGAGCATTTCAGATCCCAAGTCAGTGATCACAACCTACGAAGGAAA
ACACAAACACCAAATCCCGACCCCAAGAAGAGGTCCAGTTTAAAGATCTGCTGCAATGGC
TTCTCCTCTTCTCCAACTTCGACTACTCTGATCAACTTCCCGGCGGCGATCCACAGTT
GCTGAGCTCTCTACGCGTCCCTCTGTCCCGGCTTCTAGCCACCGTCCGTCACGCTTCTGC
AGATGCCAGACCCTGGGCAGAGCTCGTTGACCGGTTCAGCGTTTCCCGGCCACCATCGCT
CTCGGAGGCAACGTACAGTAAGGAAGAACCTTTTCTATTTCGAGCCAATTACATAAC
CTTAGTGGAATCTTACTCGCCGCGTCTCTGCTCAGCACCCCTTTCGCTCTCTTCTCCT
CGCATCGCTGGCCGCTTCTTGGCTTTTCTCTACTTTTTCCGTCGCGGCGATCAGCCGTT
GGTCATTGGAGGACGCAGTTCTCCGATCTTGAGACGCTAGGGATACTCTGCCTGTCCAC
TGTGGTGGTGATGTTTATGACCAGCGTTGGATCGCTCTTGATGTCCACTCTAGCAGTTGG
GATCATGGGCGTGGCCATCCACGGAGCGTTTCGTGCTCCCGAAGACCTGTTTCTTGAAGA
ACAAGAAGCCATTGGATCTGGACTTTTTCGCATTCTTCAACAACAATGCCTCTAATGCAGC
TGCCGCTGCCATAGCCACCTCAGCAATGTACGCGTTCGAGTCTGAGATTGTTGAAGAGA
CTACATTCCTACACCGCATTTCCAAAGTGTGATATTTATTATCATATTGAATTGTT
>G176 Amino Acid Sequence (domain in AA coordinates: 117-173,234-290)
MGSFDRQRAVPKFKTATPSPLPLSPSPYFTMPPLTPADFLDSPLLFSSNILPSPTTGT
FPAQSLNLYNNGLLIDKNEIKYEDTTPPLFLPSMVTQPLPQLDLFKSEIMSSNKTSDGY
NWRKYGQKQVKGSENPRSYFKCTYPNCLTKKKVETSLVKQMIEIVYKGSNHPKPQSTK
RSSSTAIAAHQNSSNGDGKDIGEDETEAKRWKREENVKEPRVVVQTTSDIDILDDGYRWR
KYGQKVVGKGNPNRPSYKCTFTGCFVRKHVERAFQDPKSVITTYEGKHKHIPTPRRGPV
LRSAMASPLLPSTTTPDQLPGGDPQLLSLRVLLSRVLATVRHASADARPWAELEVDRSA
FSRPPSLSEATSRVRKNFSYFRANYITLVAILLAASLLTHPFALFLLASLAASWLFYFF
RPADQLVIGGRTFSDLETGLILCLSTVVMFMTSVGSLLMSTLAVGIMGVAIHGAFRAP
EDLFLBEEQEAIGSGLFAFFNNNASNAAAAAIATSAMSRVRV*
>G174 (194..1585)
CCCAATTTGAGATTGTTTCGATTTTCGATCTACGAGATTCTTACAAGAACATAAGCAGCTTC
GGTTTTTTGGGATTATCTTATTTGGTTCGGATGATGATCTTCTCGATGTCTGTGCTAGGCT
TTGGGAATTAGATATATTTGGGGTTAAGCTCGAGTCTCTCCGGTTTTGAGTTTACTTGAG
TTTGTTAGTATTTATGGCTGAGGTGGGAAAAGTTCTGGCTAGTGATATGGAGTTAGACCA
TTCAAATGAGACTAAAGCAGTGGATGATGTTGTTGCCACTACTGATAAAGCGGAGGTCAT
ACAGTGGCTGTAACTAGAAGTGAACCGTTGTTGAAAGTTTGAATCTACTGACTGTAA
GGAGCTTGAAAAAATTTGTTCCACATACGGTAGCTTCGCAGTCGGAAGTAGATGTTGCTTC
CCCGGTATCCGAGAAAGCACCGAAGGTTTCTGAAAGTAGCGGTGCATTATCTTTGCAGTC
TGGTTCGGAAGGGAATAGTCTTTTTATTCGTGAGAAGGTTATGGAAGACGGATACAACCTG
GCGGAAATATGGACAGAACTTGTGAAAGGAAATGAGTTTGTAAGGAGCTATTACAGGTG
CACTCACCCCTAACTGCAAGCGGAAAAAACAGTTGGAACGGTCTGCGGGTGGACAAGTCGT
GGATACCGTTTACTTTGGGGAAACATGATCACCCAAAGCCTCTTGCTGGTGCTGTTCTCTAT
CAATCAGGATAAGCGAAGTGTGTTTACAGCTGTTAGTAAAGAGAAAAACATCTGGATC
CAGTGTTCAGACACTTCTGTCACAAACCGAACCAAGATCCATGGAGGATTACATGTTTC
AGTTATTCCACCAGCTGATGATGTGAAACTGATATTTACAATCAAGTAGGATAACGGG
GGACAACACTCACAAGGATTATAATAGTCTACCGCAAAGCGAAGGAAGAAAGGAGGGAA
CATTGAGCTGAGTCCAGTGGAGAGGTCAACCAATGATTACGCATTGTGGTTCACACTCA
GACTCTGTTTGATATTGTGAATGATGGTACCGATGGCGTAAATATGGTCAGAAATCAGT

AAAAGGCAGCCCATATCCAAGGAGCTACTATAGATGTTCAAGCCCTGGATGCCCCGTCAA
GAAACACGTCAGAGAGGTCTATCTCATGACACAAAGTTGCTTATAACAACCTACGAGGGAAA
ACACGACCACGATATGCCTCCAGGAAGAGTTGTTACTCATAATAACATGCTGGACTCGGA
AGTTGATGATAAAGAAGGAGATGCCAACAAGACTCCACAGAGCTCAACTCTTCAATCCAT
TACAAAAGACCAGCATGTGCAAGATCACTTAAGAAAAGAAAACGAAGACTAATGGCTTTGA
GAAAAGTCTTGATCAAGGTCCAGTTTGTGGATGAGAAGCTGAAGGAGGAAATAAAAGAGAG
ATCAGATGCAAAACAAAGATCACGCAGCCAATCACGCCAAGCCGGAAGCAAAGTCAGATGA
TAAACCCTGTTGTGTCAGAGAAGGCAGTAGGAACCCCTGGAGAGCGAGGAACAAAACC
CAAGACAGAGCCTGCCCAAAGCTAAGCATTCAAGTGTGTACCGAGTGGTAATTTATATGG
CTGTTTAAACATAGATTAGTACAGGCGATATGGTTATAGACTGTACAGTTGTTGTTTCAGG
CGGGACCAGATTAGATTAGTGTTTAATGGAATAGTATGCTTTAATACCTTTATGTAACC
ACTTCCATTGTTGTTCAAATAAGAGTTACAGGAAGAGAAGGTAACACAACAAGAGCCCTTC
TTTGTGATGGAGCCTGTGTAATAGTTGTAGCATGGGGATGTATATGATTGATTCAACC
TTATTAATGGTTATGAGACAAAACCTATC

>G174 Amino Acid Sequence (domain in AA coordinates: TBD)
MAEVGKVLASDMELDHSNETKAVDDVVATTDKAEVIPVAVTRTETVVESLESTDCKELEK
LVPHTVASQSEVDVASPVSEKAPKVSSESSGALSLSQSGSEGNPFIREKVMEDGYNWRKYG
QKLVKGNFVRSYYRCTHPNCKAKKQLERSAGGQVVDTVYFGEHDHPKPLAGAVPINQDK
RSDVFTAVSKEKTSGSSVQTLRQTEPPKIHGGLHVSVIPPADDDVKTDISQSSRITGDNTH
KDYNSTAKRRKKGGNIELSPVERSTNDSRIVVHTQTFLDIVNDGYRWRKYGQKSVKGS
YPRSYRCSSPGCPVKKHVERSSHDTKLLITTYEGKHDHDMPPGRVTVTHNMLDSEVDDK
EGDANKTPQSSTLQSIITKQHVHDLRKKTKTNGFEKSLDQGPVLDEKLKEIKERSDAN
KDHAAHNAKPEAKSDDKTTCVQEKAVGTTLESEEQKPKTEPAQS*

>G715 (1..705)
ATGGATACCAACAACCAGCAACCACCTCCCTCCGCCGCCGAATCCCTCCTCCACCACCT
GGAACCACCATCTCCGCCGAGGAGGAGGAGCTTCTTACCACCACCTTCTCCAACAACAA
CAACAACAGCTCCAACCTATTCTGGACCTACCAACGCCAAGAGATCGAACAAGTTAACGAT
TTCAAAAACCATCAGCTTCCACTAGCTAGGATAAAAAAGATCATGAAGCCGATGAAGAT
GTTGCTATGATCTCCGCGAGAAGCACCGATTCTCTTCGCGAAAGCTTGTGAGCTTTTCATT
CTCGAGCTCACGATCAGATCTTGGCTTCACGCTGAGGAGAATAAACGTCGTACGCTTCAG
AAAAACGATATCGCTGCTGCGATTACTAGGACTGATATCTTCGATTCTCTTGTGATATT
GTTCTTAGAGATGAGATTAAGGACGAAGCCGAGTCTCGGTGGTGAATGGTGGTGGCT
CCTACCGCGAGCGCGCTGCTTACTATTATCCGCCGATGGGACAACCAGCTGGTCTCTGGA
GGGATGATGATTGGGAGACCAGCTATGGATCCGAATGGTGTATGTCAGCCTCCGTCT
CAGGCGTGGCAGAGTGTGTCGAGACTTCGACGGGACGGGAGATGATGTCTTATGGT
AGTGGTGGAAAGTTCCGGTCAAGGGAATCTCGACGGCCAAGGTA

>G715 Amino Acid Sequence (domain in AA coordinates: 60-132)
MDTNQPPPPSAAGIPPPPGTTISAAGGGASYHLLQQQQQLQLFWTYQRQEIEQVND
FKNHQLPLARIKKIMKAEDVRMISAEAPILFAKACELFLELTIRSWLHAEENKRRTLQ
KNDIAAAITRTDIFDFLVDIVPRDEIKDEAAVLGGGMVAPTASGVPIYPYPPMGQAPGP
GMMIGRPAMDPNGVYVQPPSQAWQSVWQTSTGTGDDVSYGSGSSGQGNLDGQ*

>G588 (196..1599)
ATCTGAAGTGAACCAAGCTCAGGTTTTGTCTTCTCTTGTATCATTCCTTTCTCAGCAATA
TAAATTAGAGTTATATCCTTTTATAAAGGATTTTGCTTTTTCACCAACAAACCCTAAATTC
GGTGTCTCAGCAAGAATCACGTGATTCTCGTTCTCTCTCTCACGAAACCCATCATCTTC
TATCTCATTGAGAAATGGGTCAAAGTTTTGGGAGAATCAAGAAGATCGAGCGATGGTT
GAATCCACCATAGGCTCTGAAGCTTGCAGCTTTTTCATCTCAACAGCTTCAGCTTCCAAC
ACTGCCTTGTCCAAGCTTGTCTCACCAACAGTGATTCCAATCTCCAACAAGGGTTACGT
CAGTTGTTGAAGGATCTGATTGGGATTATGCTCTTTTCTGGCTAGCGTCCAACGTTAAT
AGCTCTGATGTTGTCTCTTGATCTGGGAGATGGTCATTGCCGTGTCAAAAAGGGTGCT
TCAGGTGAGGATTACTCTCAGCAAGATGAGATCAAAAGACGTGTGCTTCGCAAGCTTCAC
TTGTCTGTTGTTGTTTCAGATGAAGATCATCGTTTGGTGAAATCAGGAGCTCTTACTGAT
CTCGACATGTTTTATCTGGCTTCTTTGTACTTTTCTTTAGGTGTGATACCAATAAGTAC
GGTCTGCTGGAACCTATGTGTCTGGAAGCCTCTTGGGCTGCAGATTTGCCTAGCTGC
TTGAGTTATTATAGGGTTAGGTCTTTCTTAGCTAGGTGAGCTGGTTTTTCAGACTGTGTTG
TCTGTACCAGTGAATCTGGAGTTGTGGAGCTTGGTTCTTTAAGACATATTCAGAAGAT
AAGAGTGTGATTGAGATGGTGAAATCAGTGTGTTGGTGGGCTGACTTTGTTTCAGGCTAAA

GAAGCTCCTAAAATCTTTGGTCGACAGCTGAGTCTTGGTGGAGCAAAACCTCGGTCTATG
AGTATTAATTTCTCCCCGAAGACCGAGGATGACACGGGTTTCTCATTGGAATCGTATGAG
GTGCAAGCGATCGGAGGCTCTAATCAAGTGTATGGTTATGAGCAAGGGAAAGATGAGACA
TTGTATCTAACTGACGAGCAAAAGCCGAGGAAGAGAGGGAGAAAACAGCAAATGGAAGA
GAAGAGGCTCTAAACCATGTGTGAAGCGGAACGGCAGAGGAGGAGAAGCTGAACCAGAGA
TTCTACGCTTTGAGAGCGGTGGTGCTAACATCTCCAAGATGGACAAGGCTTCGCTCCTT
GCAGACGCAATCACTTACATCACGGATATGCAGAAGAAAATCAGGGTGTATGAAACAGAG
AAGCAGATAATGAAGAGGAGGAGAGTAATCAGATAACTCCAGCAGAGGTTGATTATCAA
CAGAGGCATGATGATGCAGTTGTAAGGCTAAGCTGTCCGTTGGAACTCATCCAGTTTCA
AAGGTGATACAAACGTTGAGGGAGAATGAAGTTATGCCTCATGATTCCAACGTGGCCATC
ACAGAGGAGGGTGTGGTTCACACATTCACCTCTCCGGCCTCAGGGTGGCTGCACCGCTGAG
CAGTTGAAGGACAAGCTTCTGCCTCTCTATCACAGTAATCACAGCAGTAACCTGCTA
TGTAATAAGTGTAACCGTGTGGAGGTTGTATCAATGTACTATTGCAAGCCAACCAAAAA
AACTCCAGCTTAGTAGGATCGTGTAAATTTTCCTTATATGTAATGTTGAGATTGTCTTT
TACATATAAAGATTTGA

>G588 Amino Acid Sequence (domain in AA coordinates: 309-376)
MGQKFENQEDRAMVESTIGSEACDFFISTASANTALSCLVSPSPSDSNLQQLRHVVEG
SDWDYALFWLASNVNSDGCVLWGDGHCVRVKKGASGEDYSQDEIKRRVLRKLHLSFVG
SDEHRLVKGALTDLDMFYLASLYFSFRCDTNKYGPAGTYVSGKPLWAADLPSCLSYYR
VRSFLARSAGFQTVLSVPVNSGVVGLSLRHIPEDKSVIEMVKS VFGGSDVQAKEAPKI
FGRQLSLGGAKPRSMSINFPKTEDDTGFSLESYEVQAIGGSNQVYGYEQGKDETLYLTD
EQKPRKRGRKPANGREELNHVEAERQRREKLNRFPYALRAVVPNISKMDKASLLADAIT
YITDMQKKIRVYETEKQIMKRRESNQITPAEVDYQQRHDDAVVRLSCPETHPVSKVIQT
LRENEVMPHDSNVAITEEGVVHTFLRPQGGCTAEQLKDKLLASLSQ*

>G1758 (69..677)
GTCCCTCCTCTTAGCTTCAACCGCCGAAAAACTAAACAACCTTCTTGGAAAAAAGAGA
AACTAAAAATGAACCTATCTTCAAACCTTAACCTAGCTCCACAGATTTCACTGAATTTT
TCAAGTTTCGATGATTTTTCAGGATACTTTTGAGAAGATCATGGAAGAAATCGGCCGTGAGG
ACCACTCGTCGTACCGACTTTGAGTTGGAGTTTCATCGGAAAAGTTAGTGGCTGCAGAAA
TCACAAGCCCCGCTTCAAACAAGCCTAGCTACCTCACCCTATGAGCTTTGAAATAGGTGACA
AAGATGAAATCAAAAAGAGGAAGAGACACAAAGAAGATCCGATTATTCACGCTCTTCAAAA
CGAAATCATCAATTGATGAAAAGGTTGCTTTAGATGATGGGTATAAATGGAGGAAATACG
GAAAGAAGCCGATAACGGGTAGTCCATTTCCAAGGCATTATCACAGTGTTTCGAGCCAG
ATTGCAACGTGAAGAAGAAGATCGAAAGAGATACGAACAATCCAGATTACATATTGACAA
CATACGAAGGTAGACATAACCAACCAAGCCCTTCTGTAGTTTATTGTGATTCAGACGACT
TTGATCTTAACCTCTCTCAACAATTGGTCTTTTCAGACGGCAAATACGTATAGTTTCTCTC
ATTCTGCTCCATATTGATCGATCGTAGTTACAAGTTTGTGTATATAGATGTATATATATA
TATCACCATTACCATCGTAATCACGTCTCACATGTAACGTACGTACATATATCTTGTTC
GGGGTTCGTTTGTAAATGTAATTGGTGGAGGTAGAATGGAAGTCATCTTGTATAGT
TGTAAGTTGTATGTAAGGTTGATAGTCATTTTTTATAAAGTAACATAATTGTACAA

>G1758 Amino Acid Sequence (domain in AA coordinates: TBD)
MNYSPNPNPSSTDFTEFFKFDDDFDTFEKIMEEIGREDHSSSPTLSWSSEKLVAAEITS
PLQTSLATSPMSFEIGDKDEIKKRKRHKEDPIIHVFKTKSSIDEKVALDDGYKWRKYGKK
PITGSPFPRHYHKCSSPCDNVKKIERDTNNPDYILTTYEGRHNHPSVSVYCDSDDFDL
NSLNNWSFQTANTYSFHSAPY*

>G2148 (66..737)
GTCTCTAATATAAGCTTGAACGTTGCTATATATAAATGTAAAGGCGAACGCATAAGAAAA
GAAAAATGGAGAATGAAGCTTTTGTAGATGGTGAATTGGAGTCTCTTTGGGGATGTTCA
ACTTTGATCAATGTTTCATCAACGAATCGAGCTTTTGCAATGCTCCAAATGAGACTGATG
TTTTCTCTTCTGATGATTTCTTCCCATTTGGTACAATTCTGCAAAGTAACATGCGGCCG
TTCTTGATGGTTCCAACCACCAAAACGAACCGAAATGTCGACTCAAGACAAGATCTGTTGA
AACCAGGAAGAAGCAAAAGTTAAGCTCGGAAAGCAATTTGGTTACCGAGCCTAAGACTG
CTTGAGAGATGGTCAAAGCCTAAGCAGTTATAATAGTTTCAGATGATGAAAAGGCTTTAG
GTTTAGTGTCTAATACATCAAAAAGCCTAAAACGCAAAGCGAAAGCCAACAGAGGGATAG
CTTCCGATCCTCAGAGCCTATACGCTAGGAAACGAAGAGAAAGGATAAACGATAGGCTAA
AGACATTGCAGAGCCTAGTTCTTAATGGGACAAAGGTCGATATAAGCACAATGCTGGAAG
ATGCTGTCCATTACGTGAAGTTCTGCGAGCTTCAAATCAAGCTCTTGAGTTCAGAAGATC

TATGGATGTATGCACCTCTTGCTCACAATGGTCTGAATATGGGACTACATCACAATCTTT
TGTCTCGGCTTATTTAAGACAAAATCATTGGAATAACATAACTTACAGTACTTGT'TTTT
TTCTCGTTCTATATTCATGATTATGGTTATTTTTTGT'TTGAGTTGTTCAATTTTCTGTC
TATTGCGTCTATGAAC'TTGACACTCTTTTGTAAATTATTATATGCTAAAGACAATTTGG
ACTAACAGCATT'TTAATAAAAAAAAAAAAA

>G2148 Amino Acid Sequence (conserved domain in AA coordinates:130-268)

MENEAFVDGELESLLGMPNFDQCSSNESSFCNAPNETDVFSSDDFFPFGTILQSNYAAVL
DGSNHQTNRNVDNRQDLLKPRKKQKLSSSNLVTPEKTAWRDGGQSLSSYNSSDDEKALGL
VSNTSKSLKRKAKANRGIASDPQSLYARKRRERINDRLKTLQSLVPNGTKVDISTMLEDA
VHYVKFLQLQIKLLSSEDLWMYAPLAHNGLMGLHNNLSRLI*

>G2379 (52..798)

CGCCGCTCACTCTCCTCCCGGTGCCGCACATTAGCAACACTACTCCCGACGAATGGAGACG
ACGACGCCGCGAGTCAAAATCAAGTGTGTCCACCGACCGCCGTTGGGAAGAGAAGACTGG
TGGAGTGAGGAAGCGACGCGACGCTGGTAGAAGCCTGGGGCAATCGTTACGTCAAGCTG
AACACCGGAAATCTCCGGCAGAATGACTGGAAAGACGTCGCCGACGCCGTTAACTCTAGA
CACGGTGATAACAGCCGTAAGAAGACCGACTTACAGTGTAAGAACCGGGTCGATACTTTG
AAGAAGAAGTACAAAACAGAGAAAGCTAAACTCTCGCCGTCGACTTGGCGTTTCTATAAC
CGCCTCGATGTTCTAATCGGTCCCCTGTGAAGAAATCGGCTGGCGGAGTTGTCAAATCA
GCGCCTTTTAAGAATCATCTGAATCCAACTGGATCGAACTCTACTGGAAGCTCTCTTGAA
GATGATGATGAGGATGATGATGAGGTTGGTGATTGGGAATTCGTTGCTAGGAAGCATCCT
CGTGTGGAAGAGGTAGATCTGAGTGAAGGATCAACGTGTAGGGAAGTACTACGGCGATT
CTCAAGTTTGGAGAAGTTTACGAGAGAATTGAAGGAAGAAGCAACAGATGATGATTGAG
TTGGAGAAGCAGAGAATGGAAGTGACAAAGGAGGTAGAGTTAAACGAATGAACATGTTG
ATGGAGATGCAGTTAGAGATTGAGAAATCAAAGCACCGGAAACCGCAAGTGCTTCAGGT
AAGAAGAACTCACATTAGG

>G2379 Amino Acid Sequence (domain in AA coordinates:19-110, 173-232)

METTTTQSKSSVSHRPPLGREDDWSEEATATLVEAWGNRYVKLNHGNLRQNDWKDVADAV
NSRHGDNRSRKKTDLQCKNRVDTLKKKYKTEKAKLSPSTWRFYNRLDVLIGPVVKKSAAGV
VKSAPFKNHLNPTGSNSTGSSLEDDDDDEVGDWEFVARKHPRVEEVDLSEGSTCRELA
TAILKFGEVYERIEGKKQMMIELEKQRMVETKEVELKRMNMLMEMQLEIEKSKHRKRAS
ASGKKNH*

>G1462 (63..1031)

CGTCGACCATCTTGCAGATTGATCTTCTCTAGATAATTTTTTTGATCGATTTAGTTTCA
TTATGGAGGACGACGACGCGAGCTTATGATCTAATCAAACACGAACTGTTATCTCAGAAG
ACGAAGTAATAATCTCACGTTATCTGAAGGGTATGGTCGTTAACGGAGATTCTTGGCCAG
ATCACTTCTATCGAAGACGCAACGTTGTTACCAAGAATCCAGATAAGGTGTTCAATTCTG
AGAGACCTAGATTCTGTGATCGTTAAACCACGAACAGAGGCTTGTGGTAAACCGATGGAT
GTGATTCCGGTTGCTGGAGGATCATTGGTCGTGATAAACTGATAAAGTCGGAGGAGACTG
GGAAGATTCTAGGTTCAAGAAGATACTCAAGTTTGCCTAAAGAGGAAACCTATAGACT
ACAAGAGAAGTTGGGTAATGGAAGAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAG
ATCATGTGATTTGCAAAATTCGGTTTATGTTTGAAGCTGAAATTAGTTTCTTGCTAAGCA
AGCATTTCTACACTACATCAGAAATCGGTTCTTGAAAATGAGCTGTTGCCATCTTATGGAT
ATTATTTATCCAATACACAAGAGGAGGATGAATTTTATCTGGACGCGATAATGACTTCGG
AAGGAAACGAGTGGCCTAGCTACGTTACCAACAACGTGTACTGTCTGCATCCATTGGAGC
TTGTGGATCTTCAAGATCGGATGTTAATGATTACGGAACCTGCATCTTCGCTAACAAGA
CTTGTGGTGAAACTGATAAATGCGATGGTGGTTACTGGAAGATCCTGCACGGTGATAAGC
TGATCAAGTCAAATTTCCGAAAGGTCATTGGTTTCAAGAAGGATTTTGAGTTCTATGAAA
CGGTGAGACAAATATATCTTTGTGATGGAGAAGAAGTGACCGTAACCTGGACTATACAAG
AGTATAGGCTTAGCAAAAACGTGAAGCAGAATAAAGTGTGTGCGTTATCAAGTTGACTT
ATGATAGATAGATATCTTACTTTGGTTTTTGTGATCATCTTAGTATCTTACGAATATTC
TAGATACACACATCTATAGGCGACCGCTCTAGACAGGCTCGTACCG

>G1462 Amino Acid Sequence (domain in AA coordinates: TBD)

MEDDDAAYDLIKHELLYSEDEVIIISRYLKGMMVNGDSWPDHFIEDANVFTKNPKVFNSE
RPRFVIVKPRTEACGKTGDCDSGCWRIIGRDKLIKSEETGKILGFKKILKFLKRPIDY
KRSWMEEYRLTNLNLWKQDHVICKIRFMFEAEISFLLSKHFTTSSESVLENELLPSYGY
YLSNTQEDEFYLDAMTSEGNEWPSYVTNNVYCLHPLELVDLQDRMFNDYGTICIFANKT
CGETDKCDGGYWKILHGDKLIKSNFQKVIQKVFYETVRQIYLCDGEEVTVTWTIQE

YRLSKNVKQNKVLCVIKLTIDR*

>G1211 (44..1120)

TGAAACCTAGATTCTGCAACTGAATTCCTAATTCGAAAAAGATGGAGGGTTCGTCGTC
GACGATAGCAAGGAAGACATGGGAACTAGAGAACAGCATTCTAACAGTAGACTCACCTGA
TTCAACCTCCGACAACATCTTCTACTACGACGATACTTCACAGACTAGGTTCAGCAAGA
GAAACCGTGGGAGAATGATCCTCACTACTTTAAACGAGTCAAGATCTCAGCGCTCGCTCT
TCTTAAGATGGTGGTTTCACGCTCGCTCTGGTGGTACAATTGAAATAATGGGTCTTATGCA
AGGTAAAGACCGATGGTGATACTATCATTGTTATGGATGCTTTTGCTTTACCAGTGGAAGG
TACTGAGACAAGGGTTAATGCTCAGGATGATGCTTATGAGTACATGGTTGAGTATTACACA
GACCAACAAGCTCGCGGGCGGCTGGAGAATGTTGTTGGATGGTATCACTCTCACCTGG
ATATGGATGCTGGCTCTCCGGTATTGATGTTTCTACGCAGACGCTTAACCAACAGCATCA
GGAGCCATTTTTAGCTGTTGTTATGATCCACAAGGACTGTTTCAGCTGGTAAGGTTGA
GATTGGTGCTTTTCAGAACATACTCTAAAGGATATAAGCCTCCAGATGAACCTGTTTCTGA
GTATCAAATACTCTTTAAATAAGATTGAGGACTTTGGTGTCACTGCAAACAGTACTA
TTCATTAGATGTCACTTATTTCAAGTCATCTCTTGATTCTCACCTTCTGGATCTACTATG
GAACAAGTACTGGGTGAACACTCTTCTCTCTCCACTGCTGGGTAATGGAGACTATGT
TGCTGGACAAATATCAGACTTAGCTGAGAAGCTTGAGCAAGCCGAGAGTCATCTGGTTCA
GTCTCGCTTTGGAGGAGTTGTGCCATCATCCCTTCATAAGAAAAAGAAGATGAGTCTCA
ACTAATAAGATAACTCGGGATAGCGCAAAGATAACTGTGGAACAGGTCCATGGACTAAT
GTCGCAGGTACATAAAGATGAATTATTCAACTCAATGCGTCAGTCCAACAACAAATCTCC
CACTGACTCGTCGGATCCAGACCTATGATTACATATTGAAGTGCTCTTCTTTTGGTTT
CTANTTTTGGATTGACCCATCATTGTTGTCTTTTCTTTTCTGTTGTGTAAAGAA
TTATAATGNCNCGCAATTTCGCGCCGCTAAAAAANACAGGAAATTGAAAAAATTCN
NCCATTCCAACATCTTTATTTAATATTATCTCCTCNATTATATAATATTCAAACATCCCT
ANTANCTTCATTGACCGTCCCCCTCCCTCCCGTGTGCTGCTGGCCCC

>G1211 Amino Acid Sequence (domain in AA coordinates: 123-179)

MEGSSSTIARKTWELNSILTVDSPDSTSDNIFYDDTSQTRFQQEKPWENDPHYFKRVK
ISALALLKMMVHARSGGTIEIMGLMQKTDGDTIIVMDAFALPVEGTETRVNAQDDAYEY
MVEYSQTNKLAGRLNVVGVYHSHPGYGCWLSGIDVSTQTLNQHQEPFLAVVIDPRTV
SAGKVEIGAFRTYSKGYKPPDEPVSEYQTIPLNKIEDFGVHCKQYYSLDVITYFKSSLD
LLDLLWNKYWNTLSSSPLLGNDYVAGQISDLAEKLEQAESHLVQSRFGGVVPSLHKK
KEDESQTKITRDSAKITVEQVHGLMSQVIKDELFNSMRQSNKSPDSSDPDMITY*

>G1048 (5..892)

GACCATGGCGGAGGAATTTGGAAGCATAGATTTACTCGGAGATGAAGATTTCTTCTTCGA
TTTCGATCCTTCAATCGTAATTGATTCTCTTCCGGCGGAGGATTTCTTCAGTCTTCACC
GGATTGATGGATCGGAGAAATCGAGAATCAATTGATGAACGATGAGAATCATCAAGAGGA
GAGTTTGTGGAATTGGATCAGCAATCGGTTTCAGATTTTCATAGCGGATCTACTCGTTGA
TTATCCAATAGCGATTCTGGCTCCGTTGATTGCGCGCTGATAAAGTTCTAACCGTTCGA
TTCTCCCGCCGCGCTGATGATTCCGGGAAGGAGAATTCGGAATTGGTTGTTGAGAAGAA
GTCTAATGATTCTGGTAGCGAGATTCATGATGATGATGACGAAGAAGGAGACGATGATGC
TGTGGCTAAAAACGAAGAAGGAGAGTAAGAAATAGAGATGCGCGGTTAGATCGAGAGA
GAGGAAGAAGGAATATGTACAAGATTTAGAGAAGAAGAGTAAGTATCTCGAAAGAGAATG
CTTGAGACTAGGACGTATGCTTGAGTGCTTCGTTGCTGAAAACAGTCTCTACGTTACTG
TTTGCAAAAGGGTAATGGCAATAATACTACCATGATGTGCAAGCAGGAGTCTGCTGTGCT
CTTGTTGGAATCCCTGCTGTTGGGTTCCCTGCTTTGGCTTCTGGGAGTAACTTCATTTG
CCTATTCCCTTATATGTCCACACAAAGTGTTCCTCCTACGTCCAGAACAGAAAAGCT
GGTTCTAAACGGGCTCGGGAGTAGTAGCAAACCGTCTTATACCGGCGTTAGTCGGAGATG
TAAGGGTTCGAGGCETAGGATGAAATACCAATCTTAACCTTGCGGCGTGACAACGCCT
TTTTTAAGTCTCTTTTGCGCATTTTGAGTTGTAGATGAGTGTCTTTAGTTTCTCTC
TCTTGTTTGTATTTCGCTGTTGAAAGTTTTCTGTCTAATATCGATAAGTTAACAGTGAA
AAAAAAAAAAAAAA

>G1048 Amino Acid Sequence (domain in AA coordinates 138-190)

MAEEFGSIDLLGDEDFDFDPSIVIDSLPAEDFLQSSPDWIGBIENQLMNDENHQEES
FVELDQQSVSDFIADLLVDYPTSDSGSVDLAADKVLTVDSPAAADDSGKENSIDLVEKKS
NDSGSEIHDDDEEGDDDAVAKRRRRVRNRDAAVRSRERKKEYVQDLEKSKYLERECL
RLGRMLECFVAENQSLRYCLQKGNNTTMSKQESAVLLLESLLGSLWLLGVNFICL
FPYMSHTKCLLRPEPEKLVNLGLSSSKPSYTGVSRRCKGSRPMKYQILTLAA*

>G986 (31..846)

CATTAAATTGGCTCCTGTGAACCTAAATTTATGGACTATGATCCCAACACCAATCCGTTTC
GACCTTCATTTCTCCGGTAACTTCCGAAAAGAGAAGTCTCGGCTTCAGCTTCTAAAGTT
GTAGAGAAGAAATGGTTAGTGAAAGATGAGAAGAGAAATATGCTACAAGATGAAATAAAC
CGGGTTAATTCGGGAGAACAGAAGCTAACCGAAATGTTAGCAAGAGTCTGTGAGAAGTAC
TATGCTCTTAATAATCTTATGGAGGAGTTGCAGAGTCGAAAGAGTCCTGAAAGTGTTAAC
TTTCAGAACAAACAGCTAACGGGGAAACGAAAACAAGAACTTGATGAGTTTGTAGCTCC
CCAATTGGACTCAGTCTCGGACCAATCGAGAACATCACCAACGATAAAGCGACGGTTTCA
ACCGCTTACTTTGCTGCTGAGAAGTCTGACACAAGCTTGACTGTGAAAGATGGATATCAA
TGGAGGAAATACGGGCAAAAGATTACGAGAGATAATCCATCTCCTAGAGCTTACTTCAGA
TGCTCGTTTTTCACCGTCTTGCTAGTCAAGAAGAAGGTGCAACGAAGTGCAGAAGATCCA
TCTTTCTTGGTAGCCACTTACGAAGGGACACATAACCACACCGGACCACATGCAAGTGTG
TCCAGGACAGTGAAACTTGATCTAGTTCAAGGTGGGCTTGAACCAAGTTGAGGAAAAGAAA
GAGAGAGGGACGATTCAAGAGGTTTGGTGCAACAAATGGCTTCTTCGTTGACCAAGAT
CCTAAGTTCACTCAGCTCTTGCGACCTGCTATTTCCGGGAGATTGATAGAGCATTCAAGA
ACATGAAAGTTCTCTAGAACATGTATATTTCTGTTTTGTTCTATTTTGTGCTCATTCCCT
AGTAAAAAGGTAAAGATTTGTTTGATCTTGATTAGGAGGCATAGATGTCAATTTTAATGT
GTGTGTATATAATTACATCAAACTAAGTATCCAAAAGGGTCACCCCCATTTTATCTTA
TG

>G986 Amino Acid Sequence {domain in AA coordinates: 146-203}

MDYDPNTNPFDLHPSGKLPKREVSASASKVVEKKWLKDEKRNMLQDEINRVNSENKKLT
EMLARVCEKYYALNMLMEELQSRKSPESVNFQNKQLTGKRKQELDEFVSSPIGLSLGPIE
NITNDKATVSTAYFAAEKSDTSLTVKDG YQWRKYGQKITRDNPSPRAYFRCSFSPSCLVK
KKVQRSAEDPSFLVATYEGTHNHTGPHASVSRTVKLDLVQGGLEPVEKKERGTIQEVLV
QQMASSSLTKDPKFTAALATAISGR LIEHSRT*

>G789 (259..1593)

GGCAAGAAGAACCTTAGCCTCTCTTTCTTCTTCTCTCTCTCTCTGTGGTACTGTT
CTGTTTCAACTTTACTCCCTCAGTTTTCAGAACAAATCCCTATCTAGAAGAGAGATAAAAC
CGAGAAGGTTTTGGAGATAGAACTTTTGTCTTCTTTTGTCCCTCCTTGCTCGATTTTT
GTTACGTGTGAAGCAATAAAAAAACTGATATAGCTAAATCTTCCATCCATTCAGAGGC
TTCTAAATCTGATCTGACATGGAACAAGTGTGCTGATTGGAATTTTGAAGATAATTTT
CACATGTCCACTAATAAAAGATCAATCAGACCAGAAGATGAATTAGTGGAGCTATTGTGG
AGAGATGGTCAAGTGGTTTTACAAAGCCAGCTCGTAGAGAACCGTCAGTCCAAGTCCAA
ACCCACAAACAAGAAACCTTAAGAAAACCAACAATATTTTCTTGACAACCAAGAAACA
GTACAAAAGCCTAAGTACGCTGCTCTAGATGATCAAGAAACCGTCTCCTGGATACAATAC
CCTCCGGATGACGTCATCGACCCCTTTCGAATCCGAGTTCTCCTCTCATTCTTCTCTTCG
ATCGATCACCTCGGAGGTCCTGAGAAGCCACGAACGATCGAAGAGACAGTTAAGCATGAG
GCTCAAGCCATGGCTCCTCCTAAGTTTAGATCCTCGGTTATAACAGTCCGACCGAGTCAT
TGCGGCAGCAACCAAGTCAACAAATATTTCATCAGGCCACTACACTTCCGGTTTCTATGAGT
GATAGAAGCAAGAAGCTCGAAGAAAGACTTGACACTTCGTCAGGTGGCTCCTCCGGTTGC
AGCTATGGAAGGAACAACAAGAAACCGTTAGTGGAAACAGTGTAACCAATTGACCGTAAA
AGAAAACATGTTATGGATGCTGATCAAGAATCTGTGTCTCAATCAGATATAGGTTTGACC
TCAACCGATGATCAAAACCATGGGTAACAAATCGAGCCAACGGTCAGGATCTACTCGAAGA
AGCCGTGCAGCTGAAGTTCATAATCTCTCAGAAAGGAGGAGAGATCGGATCAATGAA
AGAATGAAAGCTCTTCAAGAACTCATACCTCACTGCAGCAGAACAGATAAAGCTTCGATA
TTGGATGAAGCAATTGATTACTTAAAATCACTTCAAATGCAACTCCAAGTGATGTGGATG
GGAAGTGGAATGGCGGCGGCGGAGCAGCAGCAGCAAGTCCGATGATGTTTCCCGGGGTA
CAATCATCTCCATACATTAATCAGATGGCTATGCAAAGTCAGATGCAATTGTCTCAATTC
CCGGTTATGAACCGGTCCGCTCCGAGAACCATCCCGGTTTAGTATGTCAAAACCCGGTA
CAGTTGCAGCTCCAAGCACAGAACCAATCTTATCGGAGCAGCTCGCTAGGTACATGGGC
GGGATTCGCCAGATGCCGCCGGCGGGAATCAGATGCAGACCGTGCAACAACACAGCG
GACATGTTGGGATTTGGATCTCCGGCGGGACCGCAAAGTCAACTGTCCGCCACCGGCGACC
ACCGACAGTCTTCATATGGGTAAAATAGGCTGACTTGGCATATAGTTTTCTCCGAAATT
ATTCTTCTTACAGTTGGTGATGTTATTTATTTTGGTCGCCCTAAGCAAGCATAAAAGCT
AAGTCAAATGTATTATAGAGATCTAATAAGTTAGTCTCATACTTATAACTTATTTTTAAA
CAGTTGAATTATAGTATCAATCAAGTGTTGGGAACCTAAAGATCATACTGTGTCAATAC
TTTTATATTTGTTCTCAAGTTTCATCAGAAAAACAAAATAAAAAGGATAGACTAGGCCTG

CATTTGACATTATCATGGGCTTTTTTGGGTCTATGAATATGAACATTAACCCC *

>G789 Amino Acid Sequence (domain in AA coordinates: 253-313)
MEQVFADWNFEDNFMSTNKR SIRPEDELVELLWRDQVVLQSQARREPSVQVQTHKQET
LRKPNNIFLDNQETVQKPNYAALDDQETVSWIQYPPDDVIDPFESFSSHFFSSIDHLGG
PEKPRTIEETVKHEAQAMAPPKFRSSVITVGP SHCGSNQSTNIHQATTLPVMSDRSKNV
EERLDTSSGGSSGCSYGRNNKETVSGTSVTIDRKRKHVMDADQESVSQSDIGLTSTDDQT
MGNKSSQSRSGSTRRSRAAEVHNLSERRRRDRINERMKALQELIPHCSRTDKASILDEAID
YLKSLQMQLQVMWMSGMAAAAAAASPMFPGVQSSPYINQMAMQSQMQLSQFPMNRS
APQNHPLVLCQNPVQLQLQAQNQILSEQLARYMGGIPQMPPAGNQMQTVQQQPADMLGFG
SPAGPQSQLSAPATTDLSLHMGKIG*

>G2085 (1..930)
ATGTTTGGTCGCCATTTCGATTATCCCAAATAACCAGATTGGTACCGCCTCTGCTTCCGCT
GGTGAAAGACCATGTCTCTGCTCCGCTACGCTCTGGTCACATTCTTACGACGATATGGAA
GAAATCCCTCATCTGACTCTATCTATGGTGCTGCCTCCGATTTGATTCCCGATGGCTCT
CAATTGGTTGCTCACCGATCCGATGGCTCTGAATTACTTGTCTCTCGGCCACCGGAAGGG
GCGAATCAGCTTACGATCTCGTTCCGTGGACAAGTTTACGTTTTGATGCCGTTGGTGCT
GACAAGGTGGATGCTGTGTGTGTCGCTGTGGGTGGTTCTACTGAGCTTGCTCCTGGTCCG
CAGGTGATGGAACTAGCTCAACAGCAGAATCATATGCCTGTTGTAGAATATCAGAGCCGC
TGTAGCCTTCCGCAACGGGGCACAATCCTTGGATAGGTTTCGGAAGAAGAGGAATGCTAGA
TGTTTTCGAGAAGAAAGTAAGATACGGGTGTTCCGCCAAGAAGTTGCCTTAAGAATGGCACGT
AATAAAGGTCAATTCACCTCTTCAAAGATGACAGATGGGGCTTATAACTCTGGCACAGAT
CAAGATTCTGCCCCAAGATGATGCCCATCCAGAAATATCGTGTA CTACTCATTCGCGCATTAGT
TCCAAATGTACACCAATGATGCGACGTGGCCCTTCCGGCCCCAGGACTCTCTGCAATGCC
TGTGGACTTTTTTGGGCTAACAGGGGTACATTGAGGGATCTCTCAAAGAAAACAGAAGAG
AATCAGTTGGCTTTAATGAAACCGGATGATGGTGGGAGTGTGTGCTGATGCTGCTAACAAC
TTAAACACTGAAGCTGCAAGTGTGAAGAACACACTTCCATGGTTTTCTCTTGCCAATGGG
GATAATTCTAATCTGTTAGGTGATCACTAA

>G2085 Amino Acid Sequence (domain in AA coordinates: TBD)
MFGRHSIIPNNQIGTASASAGEDHVSASATSGHIPYDDMEEI PHPDSIYGAASDLIPDGS
QLVAHRSDGSELLVSRPPEGANQLTISFRGQVYVFDVAGADKVDVAVLSLLGGSTELAPGP
QVMELAQQQNHMPVVEYQSRCSLPQRAQSLDRFRKKRNARCFEKKVRYGVRQEV ALRMAR
NKGQFTSSKMTDGA YNSGTDQDSAQDDAHPEISCTHCGISSKCTPMMRRGPSGPRTL CNA
CGLFWANRGTLRDLRSKKTEENQLALMKPDDGGSVADAANNLNTEAASVEEHTSMVSLANG
DNSNLLGDH*

>G1783 (1..603)
ATGGCCCGCTTTCGCGAGTGGACAAGGGTCGATGACAAACGTTTGTAGTTAGCTCTGCTT
CAAATCCCGGAGGGTTCGCCGAATTTATAGAGAATATCGCCTATTATCTCCAGAAACCG
GTGAAGGAGGTGGAGTACTACTTCTGCGCGTTGGTCCATGATATTGAGCGGATCGAATCG
GGTAAGTATGTTTGGCCAAATACCCGGAAGACGATTACGTGAACTGACGGAAGCAGGT
GAGTCTAAGGGCAATGGGAAAAAGACGGGAATTCCTTGGTCAGAAGAGGAACAGAGGTG
TTTCTGGAAGGACTAAATAAGTTTGGGAAAGGAGACTGGAAGAACATATCGAGGTATTGT
GTGAAGTCAAGGACCTCGACGCAAGTGGCAAGCCATGCTCAGAAGTATTTTGAAGGCAA
AAGCAGGAGAGTACGAATACTAAACGCCCCGAGTATTCATGACATGACTCTGGGAGTTGCG
GTCAATGTCCCTGGATCCAACCTTGAGTCTACTGGCCAGCAACCACATTTTGGTGATCAA
ATTCCTTCGAATCAATATTATCCCTCCCAGGAAAACCTTTCGGGGTTTTGATCAGCGATGG
TGA

>G1783 Amino Acid Sequence (domain in AA coordinates: 81..129)
MAAFPQWTRVDDKRFELALLQIPEGSPNFIENIAYYLQKPVKEVEYYCALVHDIERIES
GKYVLPKYPEDDYVKLTEAGESKGNKKTGIPWSEEEQRLFLEGLNKF GKGDWKNISRYC
VKSRTSTQVASHAQKYFARQKQESTNTRKPSIHDMTLGAVNVPGSNLESTGQQPHFGDQ
IPSNQYYPSQENFRGFDRW*

>G2072 (155..793)
TCGACCCACGCGTCCGCCACGCGTCCGATCTTTTCACAGAAGACCAACCAGCTTGGCT
CGATGAGCTCCTAAGTGAGCCAGCATCACCTAAGATTAACAAAGGTCATAGACGTT CAGC
TAGTGACACAGCTGCTTACTTGAACCTAGCTTTAATGCCTTCGAAGGAAAATCATGTTGC
TGGTTCGCTTGGCAGTTCAGAACTATGATTTGTGGCAGTCCAACCTCTTATGAACAACA
CAATAAATTAGGATGGGATTTCTCTACAGCAAATGGAACATAATATCCAAAGAAATATGTC

ATGCCGAGCTTTAAATATGTCGTCGAAACCCATTGAGAAACATGTAAGCAAAATGAAAGA
AGGAACTTCTACAAAACCAGATGGTCCTAGATCAAAGACTGACTCAAAACGTATCAAACA
TCAAAATGCTCATCGAGCGCGTTTGAGAAGGCTTGAGTACATATCAGACCTTGAAAGGAC
CATCCAAGTGCTACAAGTTGAAGGATGTGAAATGTCATCTGCCATTCACTACTTGGATCA
GCAGTTACTCATGCTTAGCATGGAAAATAGAGCTTTAAACAACGTATGGATAGTTTAGC
AGAAATCCAAAAGCTTAAACATGTGGAGCAGCAATTGCTTGAGAGAGAGATAGGAAACCT
ACAGTTTCGACGACGACCAACAACAACCACAGCAAAACCAAAAACAAGTCCAAGCAATACA
AAATCGATACACCAATATCAACCACCTGTTACACAAGAACCCGATGCCAATTTGCAGC
CTTGGCAATATGATTAGGAAATATGGATACATTGTTTCAGATTAAGCTGAGCTCCTCTTG
CTCTACCTTAATGTCCATACAACATAGGTGAACCTGATGTTTGTAGCCTTGAATGAAAAC
CTAAAAAGCATCGTTATGTAAATCAAAATGTGGTTGCCCATATCCTCCTCTATTTGCATT
TCTCTCTATTTATGGCATGGTAGAGAATCTCTTGTCAGAACTTCATGTTATGTAATAA
CTTGTAATCCTCTTATTTTCATCTATTATATATATGAATAAGTAATTTTTTTGCCAAAAA
AAAAAAAAAAAAAAAAAAAA

>G2072 Amino Acid Sequence (conserved domain in AA coordinates:90-149)

MPSKENHVGSSWQFQNYDLWQNSYBQHNLGWDFSTANGTNIQRNMSCGALNMSSKPI
EKHVS KMKEGTSTKPDGPRSKTDSKRIKHQNAHRRLRRLEYISDLERTIQVLQVEGCEM
SSAIHYLDQQLMLSMENRALKQRMDSLAEIQKLKHVEQQLLREIGNLQFRRHQQPQO
NQKQVQAIQNRVTKYQPPVTQEPDAQFAALAI*

>G931 (85..1071)

GGAGGTTCTTTGACAGACACATGTATCATCAATCTTCTCTGTTGAAGCAGAGAGAGAG
AGCTAATTGTTGCCCTCTGAGTCACATGGATAAGAAAGTTTCATTTACTAGCTCTGTGGCA
CATTTCAACTCCACCATACCTTAGTACTTCCATCTCATGGGGACTTCCAACCAATCCAAT
GGTGTGACTGAATCACTGAGTTTGAAGGTGGTAGATGCAAGACCAGAACGCTTTATAAAC
ACAAAGAAATATCAGTTTCCAGGACCAGGATTCATCTTCAACTCTGTCTCTGCTCAATCT
TCTAACGATGTTACAAGTAGTGGAGATGATAACCCCTCAAGACAAATCTCATTTTGTAGCA
CATTGAGATGTTTGTAAAGGATTGAAGAACTCAAAGGAAGCGATTGCAATTAAATCA
GGCTCCTCCACGGCAGGAATCGCTGATATTCACTCTTCTCCTTCCAAGGCTAAGTTCTCA
TTTCACTATGCCGATCCACATTTTGGTGGTTAATGCCTGCGGCTTACCTACCACAGGCA
ACAATATGGAATCCCCAAATGACTCGAGTTCGCTACCATTCCGATCTCATAGAGAATGAG
CCTGTCTTTGTCAATGCAAGCAATTCCATGCAATTATGAGGAGGAGGCAACAGCGTGCT
AAGCTAGAGGCGCAAAACAACTAATCAAAGCCCGTAAGCCGTATCTTCATGAATCTCGA
CATGTTACAGCTCTTAAACGACCTAGAGGATCTGGTGGAAAGATTCCTAAACACCAAAAAG
CTTCAAGAATCTACAGATCCAAAACAAGACATGCCAATCCAACAGCAACACGCAACGGGA
AACATGTCAAGATTGTGCTTTTATCAGTTGCGAAGACAGCAATGACTGTGATTGTTCAACC
ACTTCTCGCTCTGACATCACATCTGCTTCTGACAGCGTTAATCTCTTTGGACACTCTGAA
TTTCTGATATCAGATTGCCCATCTCAGACAAACCAACAATGTATGTTTCATGGTCAATCA
AATGACATGCATGGAGGTAGGAACACACACCATTCTCTGTCCATATCTGAGCCGCTGGA
ATCTGGTAAATGTGTACCTTCTACAAAAAAGGGAAAGTCATCCTTGGCTGCTACTTCGCT
TATTAGCTAGTTCTTATTTACACGCTTTGTCCAGATATC

>G931 Amino Acid Sequence (domain in AA coordinates: TBD)

MDKKVSFTSSVAHSTPPYLSTSIWGLPTKSNVTELSLKVVDARPERLINTKNISFQD
QDSSSTLSSAQSSNDVTSSGDDNPSRQISFLAHSVDCKGFEETQKRFAIKSGSSTAGIA
DIHSSPSKANFSFHYADPHFGGLMPAAYLPQATIWNPMQTRVPLPFDLIENEPVFNKQ
FHAIMRRRQQRAKLEAQNKLKARKPYLHESRHVHALKRPGRSGGRFLNTKKLQESTDPK
QDMPQQQHATGNMSRFVLYQLQNSNDCDCSTTSRSDITSASDSVNLFGHSEFLISDCPS
QTNPTMYVHGQSNDMHGGRNTHHFSVHI*

>G278 (93..1874)

TCGATCTTTAACCAAAATCCAGTTGATAAGGTCTCTTCTGTTGATTAGCAGAGATCTCTTTA
ATTTGTGAATTTCAATTCATCGGAACCTGTTGATGGACACCACCATTGATGGATTCCGCCG
ATTCTTATGAAATCAGCAGCACTAGTTTCGTGCTACCGATAACACCGACTCCTCTATTG
TTTATCTGGCCGCCGAACAAGTACTACCGGACCTGATGTATCTGCTCTGCAATTGCTCT
CCAACAGCTTCGAATCCGTCTTTGACTCGCCGGATGATTTCTACAGCGACGCTAAGCTTG
TTCTCTCCGACGCGCGGGAAGTTTCTTCCACCGGTGCGTTTTGTGTCAGGAGAAGCTCTT
TCTTCAAGAGCGCTTTAGCCGCGCTAAGAAGGAGAAAGACTCCAACAACACCGCCGCCG
TGAAGCTCGAGCTTAAAGGAGATTGCCAAGGATTACGAAGTCGGTTTCGATTTCGGTTGTGA
CTGTTTTGGCTTATGTTTACAGCAGCAGAGTGAGACCGCCGCCCTAAAGGAGTTTCTGAAT

CGCGAGACGAGAATTGCTGCCACGTGGCTTGCCGGCCGGCGGTGGATTTCATGTTGGAGG
 TTCTCTATTTGGCTTTTCATCTTCAAGATCCCTGAATTAATTACTCTCTATCAGAGGCACT
 TATTGGACGTTGTAGACAAAGTTGTTATAGAGGACACATTGGTTATACTCAAGCTTGCTA
 ATATATGTGGTAAAGCTTGTATGAAGCTATTGGATAGATGTAAAGAGATTATTGTCAAGT
 CTAATGTAGATATGGTTAGTCTTGAAAAGTCATTGCCGGAAGAGCTTGTTAAAGAGATAA
 TTGATAGACGTAAAGAGCTTGGTTTGGAGGTACCTAAAGTAAAGAAACATGTCTCGAATG
 TACATAAGGCACCTTGACTCGGATGATATTGAGTTAGTCAAGTTGCTTTTGAAGAGGATC
 ACACCAATCTAGATGATGCGTGTGCTCTTCATTTGCTGTTGCATATTGCAATGTGAAGA
 CCGCAACAGATCTTTTAAAACCTTGATCTTGCCGATGTCAACCATAGGAATCCGAGGGGAT
 ATACGGTGCTTCATGTTGCTGCGGATGCGGAAGGAGCCACAATTGATACTATCTCTATTGG
 AAAAAGGTGCAAGTGCATCAGAAGCAACTTTGGAAGGTAGAACCGCACTCATGATCGCAA
 AACAAGCCACTATGGCGGTTGAATGTAATAATATCCCGGAGCAATGCAAGCATTCTCTCA
 AAGGCCGACTATGTGTAGAAATAC TAGAGCAAGAAGACAAACGAGAACAAATTCCTAGAG
 ATGTTCTCTCCCTCTTTTGCAGTGGCGGCGGATGAATTGAAGATGACGCTGCTCGATCTTG
 AAAATAGAGTTGCACCTTGCTCAACGCTCTTTTCCAACGGAAGCACAAAGCTGCAATGGAGA
 TCGCCGAAATGAAGGGAACATGTGAGTTTATAGTACTAGCTCGAGCCTGACCGTCTCA
 CTGGTACGAAGAGAACATCACCGGTGTAAAGATAGCACCTTTCAGAATCCTAGAAGAGC
 ATCAAAGTAGACTTAAAGCGCTTTTCTAAAACCGTGGAACCTCGGGAACGATTCTTCCCGC
 GCTGTTCCGCGAGTGTCTGACCAAGATTATGAACGTGTGAGGACTTGACTCAACTGGCTTGCG
 GAGAAGACGACACTGCTGAGAAACGACTACAAAAGAAGCAAAGGTACATGGAAATACAAG
 AGACACTAAAGAAGGCTTTAGTGTAGGACAATTTGGAATTAGGAAATTCGTCCCTGACAG
 ATTCTGACTTCTTCCACATCGAAATCAACCGGTGGAAGAGGTCTAACCGTAAACTCTCTC
 ATCGTCTGTCGCTGAGACTCTTGCTCTTAGTGTAAATTTTGTGTACCATATAATTCTGT
 TTTCTATGATGACTGTAACTGTTTATGTCTATCGTTGGCGTCATATAGTTTCGCTCTTCGT
 TTTGTCATCTGTGTATTATTGCTGCAGGTGTGCTTCAAACAAATGTTGTAACAATTTGAA
 CCAATGGTATACAGATTGTAAATATATATTTATGTACATCAACAATAAAAAAAAAAAAAA
 AAAA

>G278 Amino Acid Sequence (domain in AA coordinates: 2-593)
 MDTTIDGFADSYEISSTSFVATDNTDSSIVYLAEEQVLTGPDVLSALQLLSNSFESVFDSP
 DDFYSDAKLVLSDGREVSFHRCLVSARSSFFKSALAAKKEKDSNNTAAVKLELKEIAKD
 YEVGFDVSVTVLAYVYSSRVPPKGVSECADENCCHVACRPVDFMLEVLYLAFIFKIP
 ELITLYQRHLLDVVDKVVIEDTLVILKLANICGKACMKLLDRCKEIIVKSNDVMVSLEKS
 LPEELVKEIIDRRKELGLEVPKVKHVS NVHKALDSDDIELVKLLKEDHTNLDACALH
 FAVAYCNVKTATDLLKLDLADVNHRNPRGYTVLHVAAMRKEPQLILSLLEKGASASEATL
 EGR TALMI AKQATMAVECNNIPEQCKHSLKGRLCVEILEQEDKREQIPRDVPPSFAVAAD
 ELKMTLLDLENRVALAQRLFPTAQAAEIAEMKGTCEFI VTSLEPDRLTGKRTSPGVK
 IAPFRILEEHQSRKALSKTVELGKRFFPRCSAVLDQIMNCEDLTQLACGEDDTAEKRLQ
 KKQRYMBIQETLKKAFSEDNLELGNSSSLDTSSTSKSTGGKRSNRKLSHRRR*

>G2421 (1..630)
 ATGGAGGGTTCGTCCAAAGGGTTGAGGAAAGGTGCATGGACTGCTGAAGAAGATAGTCTC
 TTGAGGCAGTGATTGGTAAGTATGGAGAAGGCAAATGGCATCAAGTTCCTTTAAGAGCT
 GGGCTAAATCGGTGCAGGAAAAGTTGTAGACTAAGATGGTTAAACTATTTGAAGCCAAGT
 ATCAAGAGAGGAAAATTTAGTTCTGATGAAGTTGATCTTCTTCTTCGTCTTCATAAGCTT
 CTAGGAAATAGGTGGTCCTTGATTGCTGGTCGATTACCTGGTCGGACCGCTAATGATGTC
 AAGAACTACTGGAACACCCATCTGAGTAAGAAGCATGAACCGTGTGTAAACTAAGATA
 AAAAGGATAAATATTATAACCCCTCCTAATACACCGGCCCAAAAGTTTGTGAAAATAGT
 ATCACATGTAACAAAGATGATGAGAAAGATGATTTTGTGGATAATTTTATGGTTGGAGAT
 AATATATGGTTGGAGCGTTTGCTAGACGAGGGCCAAGAGGTAGATGTGCTGGTTACAGAA
 GCGGCGGCAACAGAAAAGGAGGGGCACTTTGGCGTTTGACGTTGAGCAACTTTGGAATTTG
 TTCGATGGAGAGACTGTGATCTTTGATTAGTGTTTATAAACGTTTGTGTTCTCTGTTTTG
 TGAGGTTTCTCTATTTAATTTAGTATCTATTTTCTAAATTAATAATATCTTATAGTATT
 TTAGGCAAAACCTTATGTTTCCGTTTCTGTCGCGCCGCTCTAG

>G2421 Amino Acid Sequence (domain in AA coordinates: 9-110)
 MEGSSKGLRKGAWEEDSLLRQCIGKYEGKWHQVPLRAGLNRCKRKSRLRWLNLYLKPS
 IKRGKFSSEVDLLRLHLKLLGNRWSLIAGRLPGRANDVKNYWNTHLSKKHEPCCKTKI
 KRINIITPNTPAQKVCENSITCNKDDKDDFDVDFNMVGDNIWLERLLDEGQEVDFVLVTE
 AAATEKEGTLAFDVEQLWNLFDGETVIFD*

>G2032 (53..1789)

TCCTCCACAGAGTAAGAACTTCCATACTTTGCTCTAGATTTCCTTGAGAAAAGATGCAGCC
GATCTTCCATGCGATCCTTAAAAATGACCTTCCAGCTTTTTTAGAGTTGGTAGAAGATAG
TGAATCGTCTCTGGAGGAGAGAAACGAGGAAGAACTTGAACAAACACGGTTTTGCACAT
GGCTGCAAAGTTTTGGTCACCGAGAACTCGTCTCCAAGATTATTAGACTCCGACCTTCCCT
CGTGTCTTCCCGCAACGCATACAGAAACACCTTTGCATCTTGCTGTATCTCTGGAGA
TGTAACCATAGTTATGTCAGATGTTAGAGACTGGATTGGAAGTGTGTTCTGCGACCAATAT
CAACAACCAACACACCACTCCACTTGGCTTGCCGTAGCAATTCCATAGAGGCTGCCAGACT
CATCGCGGAAAAGACACAATCAATTGGCCTCGGTGAACCTATTCTCGCCATATCAAGTGG
ATCCACTAGTATCGTAGGGACTATACTGGAGAGATTCCAGACCTAGCTAGGGAAGAAGC
TTGGGTGTTTGAAGACGGCTCACAATCAACGCTACTGCATCATCGTGTGATAAGGGGAGA
CTTTGAACGTGACAACATATATTGTAGGGCTCGATCAAGGATTAGAAGAAGCACTTAACCC
CAATGGTTTTATCACCCTTGTCATCTTGCGGTCTCAGAGGCTCGGTGTGATCCTGGAGGA
GTTCTTGGACAAGGTTCCATTGTCTTTTACGCTCAATTACGCCGTCGAAAGAGACAGTCTT
TCATCTCGCTGCTCGAAACAAAAATATGGATGCCTTTGTTTTTATGGCAGAGAGTTTGGG
AATTAACAGCCAAATTCTTCTACAGCAAACCGATGAAAGTGGCAACACTGTCTTACATAT
TGCTGCATCCGTCTCTTTTGATGCTCCTCTTATACGTTACATTGTTGGTAAGAATATAGT
AGATATCACGTCGAAAGAACAAAGATGGGTTTTGAAGCTTTTCAACTTCTCCCTCGAGAAGC
CAAGACTTTGAGTTGTTATCAAGTGGCTGAGATTGGTACCGAGCTTCCAGAGAGCT
GGATTCTGAGAACAAATGTAGAAACAACACGAAGGCTCTCAAGAGGTCGAGGTAATACGGTT
GCTAAGGATTATAGGAATAAAACACATCAGAGATAGCAGAGAGAAAGAGAAGCAAGGAACA
GGAAGTGGAAAGAGGTCGTGAGAACTTGGAAATACAGATGCATATAGAAGCATTACAGAA
TGCAAGAAATACGATTGCTATAGTGGCAGTCTTGATTGCTTCAGTTGCTTATGCCGGTGG
GATAAACCCCTCCGGGGGGCGTCTACCAAGACGGCCATGGAGAGGGAAATCCTTAGTGGG
GAAAACAACGGCGTTTTAAGGTTCTTTCGCATATGCAACAACATCGCATGTTCCAGTCCTT
GGGCATCGTTATTCTTTCTGTTAGCATATACCTTACAAGAGGAAACCCCTTAAAGCAGATT
ATTGGTGGCCACGCATAGGATGATGTGGGTTTCTGTAGGTTTCATGGCGACGGCTTATAT
AGCGGCGCTTTGGGTGACCATACCGCATTATCATGGAACACAATGGTTATTTCCAGCAAT
TGTAGCCGTTGCTGGTGGAGCGTTGACCGTACTCTTTTCTATCTCGGAGTTGAGACCAT
CGGTTCATTGTTTTAAGAAGATGAATCGTGTAGGGGATAATATACCTTCCCTTTGCAAGAAC
CAGTTCAGATTTAGCCGCTTCCGGAATAACAGGCTATTTCACCTATTAAGAAAACTGGT
TTTTCAATTTCCCTGTAACCTGTGTAATTGTGTATGTG

```
>G2032 Amino Acid Sequence (domain in AA coordinates: entire protein)
```

MQPIFHAILKNDLPAPFLELVEDSESSLEERNEEEHLNNTVLHMAAKFGHRELVSKIIELR
PSLVSSRNAYRNTPLHLAAIIGDENVIMQMLETGLEVCSARNINNHTPLHLACRSNSIEA
ARLIAEKTQSIGLGELILAISSGSTSIVGTILERFPDLAREEAWVVEDGSQSTLLHHACD
KGFELTTILLGLDQGLEEALNPNGLSPLHLAVLRGSVVILEEFLDKVPLSFSSITPSKE
TVFHLAARNKNMDAFVFMAESLGINSQILLQQTDESGNTVLHIAASVSFDAPLIRYIVGK
NIVDITSKNKMGFEAFQLLPREAQDFELLSRWLRFGTETSQELDSENNVEQHEGSQVEV
IRLLRIIGINTSEIAERKRSKEQEVERGRQNLEYQMHI EALQNARNTIAIVAVLIASVAY
AGGINPPGGVYQDGPVRGMSLVGKTTAFKVFACINNALFSTLSGIVILLVSIIPYKRKPL
KRLLVATHRMWVSVWGFMATYIAASWVTIPIHYGHTQWLFPAIVAVAGGALTVLFFYLGV
ETIGHWEFKMNRVGDNIPSFARTSSDLAVSGSKGYFTY*

>G1396 (83..313)

TCGACCTCGTTTCCTTTCCCTCCTCTCTTCCCTACCATTAGTACGTTACTGGAGCTGATCTC
ACGTATATTTTGGATCGTAATCATGGACGGCGAAGATTTTGC CGGAAAGGCGGCTGCTGA
AGCCAAGGGATTGAACCCGGGATTAAATCGTGCTGCTTGTTGTTGGAGGTCGCTTCTTGT
GTTCTTAATCGCCAACTACGTGCTTTACGTTTATGCTCAGAAGAACCTACCTCCAAGGAA
GAAGAAGCCCGTTTCCAAAAAGAGCTCAAGCGGGAGAAGCTAAAGCAAGGAGTCCCTGT
CCCTGGAGATAAAAAGCCAGCTTAAGCTTCCCTTCACTTGTGCTCCTTCAAAGCGGTTTT
TGTTTCGGTTACCAAAATTTACCCCTTGCGGGTTTTTTTCTTCCCTTACTTCTGTCAAGAG
ATTACTTTTGAGCCT

>G1396 Amino Acid Sequence (domain in AA coordinates: TBD)

MDGEDFAGKAAAEAKGLNPGLIVLLVVGGLLVFLIANYVLVYVAQKNLPPRKKKPVSKK
KLKREKLKOGVPPVPE*

>G619 (382..2748)

ATTTTTTCCAATCTGCAAATTTTAGTCTATGTCTGTTCCCTTGCTCCCTCTTCTCAGT

A
 AACTGCAAAATGGAGGAAGAAGAAATCCTTCTCTGAAACCCTTGTTCATTGATTCTCTC
 TTCTCTCTCTCTCTCTCTCTCTGTCTCTGATTCTGTTATCCACACTTATGACTCATCTT
 TCCCGTCAATAGCTAAGTTTGCCTCTTCTTTGTGAAATTTAGCTGAAAAAGGAGAGGAAT
 TCCGAATTCTGTCACTTCAAAGCTCGAATTTTGCAAACTTTCTCTTTGATGGGTTTTACTT
 GTTTTGTGTAATCTGATTAAAAATAGAAACTTTTGTTTTCTCTTGTCTCCTTTTGCT
 CTAAAAAGAGAAGCTTTTCAATAGGAATTTGACTGTAATCTGAGATTGCGGAGTGGAA
 GAGGAGGAGAATGATGATGTAGGAGTAGGAGTAGGAGGAGGAACAAGAATTGACAAGGGT
 AGCGTTGGAATTTACCATCTTCTTCTTCTTCTCATGCTCTTCCGGATCATCATCGTCATCA
 TCTTCTACAGGCTCTGCATCTTCCATTTACTCTGAGCTTTGGCATGCTTGTGCTGGTCTC
 CTCACTTGTCTTCCCAAGAAAGGCAATGTAGTTGTCTATTTCCCTCAAGGTCATTTGGAC
 CAAGATGCTATGGTTTCATATTCTGCTCTCTCTTGAAATCCCCAAATTTGACCTTAATCCC
 CAAATCGTCTGCAAGGTTGGTTAATGTCCAGTTGCTTGCTAATAAGGACACCGATGAGGTC
 TACACTCAAGTCACTCTGCTTCCACTTCAAGAGTTTTCGATGCTAAATGGGAGGGGGA
 GAGGTCGAAGGAGTTAGGAGGGGAGGAAGAGAGGAACCGAAGCTCATCCGTCGAAGCGACA
 CCTCATATGTTCTGTAATAAACCCTTAACAGCGCTCTGACACAAGCACATGGAGGCTTCTCT
 GTACCTAGAAGAGCCGCTGAAGATTGTTTGTCTCTCTTGACTACAAACAACAGAGGCCA
 TCTCAAGAGCTCATTGCAAAGGACCTCCATGGAGTAGAGTGGAGTTTCCGCATATCTAT
 AGAGGTCAACCAAGGAGGCATCTACTACCACTGGTTGGAGTATCTTTGTGAGTCAAAAG
 AATCTCGTCTCTGGTGATGCGGTTCTCTTCTGAGAGACGAAGGAGGAGAGCTGAGATTA
 GGAATCAGAAGAGCAGCAGCGGCAAGAAATGGACTTCTTGACTCAATCATTGAGAAGAAAT
 TCATGTTCAACACTCTGTCTCTTGTGGTCAATGCTGTATCTACAAAAGCATTTTCAT
 GTGTTTCAAGCTTCCAGGAGCGACGATGCAAGAGTTTGTGATTCCTTATGAGAAGTATATC
 ACAAGCATCAGGAGTCTGTTTGCATAGGCACAAGATTAGAATGCGATTTGAAATGGAC
 GATTCTCTGAGAGAAGATGCGCTGGTGTAGTGACTGGAGTCTGTGACTTGGACCCGAT
 AGGTGGCCAAACTCTAAATGGAGGTGCTTGTGGTGCGATGGGATGAGTCTTTGTGAGT
 GATCATCAAGAAAGAGTTTACCTTGGGAGATTGATCCCTCGGTTTCTCTCCACACTTG
 AGCATTCACTCATCTCCAAGGCTAAAAGGCCATTGGGACAGGTTTACTGGATACTACCCCA
 CCGGAAACCCCAATAACAAAAGGGGTGTTTTTGGACTTTGAGAGTCGGTTAGACCC
 TCTAAGGTTTTCGAAGGTCAAGAAATATAGGTTTCTGCATCACCCCTACAGGGGTTTGTG
 GTTATGAACCGCGGATACTGGATTTTGGCATGCAGTCTCATGCAATCCAGTCTTGTG
 TCGAGTAGAGTCAAGGATCGATTTGGTGAGTTTGTAGATGCTACTGGCGTGAAACCCAGCT
 TGTTCAAGGTGTTATGGACCTGGATAGGTTTCCAAGGCTTTCGAAGGTCAAGAAATTTGC
 TCGCTTAAATCATTCCCGCAATTTGCTGGTTTCAGTCCAGCTGCTGCTCCTAATCCCTTT
 GCTTACCAAGCCACAAGTCAAGTTACTATCCGCTAGCTTTGCATGGGATTAGGAGCAC
 CATGTTTCCGATCTAGAATCCATACAATCGGGAAACCAATCTCGGTCCTCCCTTACGTT
 GCAATAAATTTTGTGGAAGAGACTAGAAAGTTTGATGACAAAATGAAGGTGGCTTACCA
 AATAATGTTTACAGCTGATTGTCCATTCAAGATTGATATGATGGGAAAACAGAAAGGCAGT
 GAGTTGAATATGAATGCTTCATCAGGATGTAACTTTTTCGATTCTCCTTACCAGTGGAG
 ACACCTGCATCTAAGCCGCAAGCTCGAGCAAAAGAATCTGTACAAAGGTTCAAGCAA
 GGAAGCCAAGTGGGAGAGCTATTGATTTGTGCGGACTTAACGGGTATGATGATCTCCTT
 ATGGAGCTTGAACGGCTGTGTTCAACATGGAAGGGCTTCTCAGGGATCCTGAAAAGGATGG
 AGGATCTTATATACTGATGTGAGAACGATATGATGTCGTTGGCGATGATCCATGGCAT
 GATTTCTGCAATGTGGTGTGGAAGATACATCTATACACAAAGAGAGATGGAGAATGCG
 AATGACGATAACAAGAGTTGTTTATAGCAAGCTGCTCTCATGATGGAAGCATCAAAGTCA
 TCTTCTGTGAGCCAGCCTGATTCTTCTCTTACAATCACTAGGGTTTGATACCCATAAAGA
 AGCTTATTTCTATGTTTTTAAAGTGTGTTTTGCTCACAAAGAAGCTTCAACTTTATCTTT
 GTCTTTGAATCCATTATGTGTTTGTGTTGTGTTTCTTCTGGTCTCCATGGATGTCTCATG
 TGTACCGTTTTACTCGAGAGATATGTGAGTTTATGGGATGTGTAAAGCATGCCATTGGAT
 TTTAAGGTTTTCAAAATTACAATATATATATATTAGTTTTGAAGTTAAAAA

>G619 Amino Acid Sequence (domain in AA coordinates: 64-406)
MEFDLNTIEIAEVEEEEENDDVGVGVGGGTRIDKGRLGISPPSSSSSCSSGSSSSSSSTGSAS
SIYSELWHACAGPLTCLPKKGNVVVYFPQGHLEQDAMVSYSSPLBIPKFDLNPQIVCRVV
NVQLLANKDTEVYTVQVTLPLQEFMSMLNGEGKEVKELGGEEERNGSSSVKTRPHMFCKT
LTASDTSSTHGGSFVPRRAEDCFAPLDDYKQQRPSQELIAKDLHGVEWKFRHIYRGQPRRH
LTTGWSIFVSQKNLVSGDAVFLRDEGGEELRLGIRRAAPCNGTLPDSIIEKNSCSNLS
LVNAVSTKSMFHVYSPRATHAEFVIPYELKYITSIRSPVIGTLPFRMFEEKNSDPERRC

AGVVTGVCDDLDPYRWPNSKWRCLLVRWDES FVSDHQERVSPWEIDPSVSLPHLSIQSSPR
PKRPWAGLLDTPPGNPITKRGGFLDFEESVRPSKVLQGGQENIGSASPSQGFDMNRRIL
DFAMQSHANPVLVSSRVKDRFGEFVDATGVNPACSGVMDLDRFPRVLQGGQEICSLKSFQ
FAGFSPAAPNPFAYQANKSSYYPLALHGIRSTHVYPYQNPYNAGNQSSGPPSRAINFGEE
TRKFDAQNEGGLPNNVTADLPFKIDMMGKQKGSELNMNASSGCKLFGFSLPVETPASKPQ
SSSKRICTKVHKQGSQVGRAIDLRLNGYDDLMELERLFNMEGLLRDPEKGWRILYTDS
ENDMMVVGDDPWHDFCNVVKIHLTYKKEEVENANDDNKSCLEQAALMMEASKSSSVSQPD
SSPTITRV*

>G2295 (33..917)

GTAATATATAACAATAACTCAGGTTACAAAGGATGGTTCGGAAAGTGGTCGACCTACAAA
GGATAGCGAACGATAAGACAAGGATAACAACCTTACAAGAAGAGGAAAGCTAGTCTTTTACA
AGAAGGCACAAGAGTTCTCAACTCTCTGCGGCGTCGAGACATGTCTCATCGTCTACGGTC
CCACGAAGGCTACCGATGTGGTGATTTCCGAGCCAGAGATATGGCCGAAGGACGAGACCA
AAGTCAGGGCCATCATACGCAAGTACAAAGACACAGTGTGACCCAGCTGCAGGAAAGAAA
CCAACGTGGAGACTTTCGTCAACGATGTAGGGAAAGGAAACGAGGTGGTGACTAAAAAGA
GAGTGAAGCGTGAGAATAAGTATTCTAGTTGGGAGGAGAAGCTAGACAAGTGTTCACGAG
AGCAACTACATGGGATTTTCTGTGCGGTGGATAGCAAGTTAAATGAAGCTGTAACGAGAC
AGGAGCGTAGTATGTTTAGGGTTAATCATCAAGCCATGGACACACCATTCCCGCAGAATT
TAATGGACCAACAATTTCATGCCACAGTATTTTCATGAGCAGCCACAGTTTCAAGGCTTCC
CTAATAATTTCAATAATATGGGTTTCTCGTTGATTTACCTCATGATGGTCAGATTCAAA
TGGACCCAAATCTCATGGGAGAAGTGGACCGACTTGGCTTTGACTCAAAGCTTGATGATGT
CAAAGGGAACGATGGTACTCAATTCATGCAGAGGCAAGAACAACCATATAATCGTG
AACAGGTTGTATCGAGGTCTGCAGGTTTCAATGTAAACCGTTTATGGGATATCAAGTCC
CGTTTAATATCTCTAATTGGAGATTATCGGGAATCAAGTTGAAAATTGGGAGCTTTCAG
GGAAGAAAACGATATGATTTGAATTACGGAGCTTTATTAGTTTTTAGGGTTTATAGTTT
TG

>G2295 Amino Acid Sequence (domain in AA coordinates: TBD)

MVPKVVDLQRIANDKTRITTYKKRKASLYKKAQEFSTLCGVETCLIVYGPTKATDVVISE
PEIWPKDETKVRAIRKYKDTVSTSCRKETNVETFVNDVGKNEVVTKKRVKRENYSSW
EEKLDKCSREQLHGIFCAVDSKLNCAVTRQERSMFRVNHQAMDTPFPQNLMDQQFMPQYF
HEQPQFQGFPPNNFNNMGFSLISPHDGGIQQMDPNLMEKWTDLALTQSLMMSKGNDDGTQFMQ
RQEOPYYNREQVVSRSAGFNVNPFMGYQVPFNI PNWRLSGNQVENWELSGKTTI*

>G312 (1..1755)

ATGGCTTACATGTGCACTGATAGTGGCAATCTAATGGCTATTGCTCAACAAGTCATCAAA
CAGAAGCAGCAACAAGAACAACAACAGCAGCAACATCATCAAGACCATCAGATTTTGGT
ATTAATCCTTTGTCTCTTAACCCATGGCCCAATACTTCCCTCGGGTTTGGGCTTTAGGT
TCGGCTTTTCCCGACCCGTTTCAAGTTACCGGCGGCGGAGATTCCAACGATCCTGGCTTT
CCTTTTCTTAACCTTAGACCACCAACACGCAACAACACCGGCGGTGGGTTCAAGTTATCT
GATTTCCGGCGGTGGAACCGCGGCGCGGAGTTTGAGTCCGACGAGTGGATGGAGACTCTT
ATCAGCGGTGGAGACTCCGTTGCAGACGGTCTGATTGTGACACCTGGCATGATAATCCC
GATTACGTAATCTACGGTCTCTGATCCATTGATACTTACCGAGTCGACTCAGTGTCCAA
CCGTGAGATCTAAACCGAGTCATTGACACGTCGAGTCCGCTTCCCTCCGCGACCTTGTGG
CCTCCTTCTTCGCCATTATCGATTCTCCGCTTACTCATGAGTCACCAACCAAGAAGAT
CCAGAGACTAACGACTCCGAAGACGATGACTTCGACCTAGAACCACCTCTCCTCAAAGCT
ATATACGACTGTGCACGGATCTCAGACTCTGACCCTAACGAAGCTTCCAAGACGCTTCTT
CAGATCCGAGAATCTGTATCGGAGCTAGGTGATCCGACGAGCGAGTTGCATTTTACTTC
ACGGAAGCTCTCTCCAACAGACTGTCTCCTAATTCCGCCGGCGACGTCGTCTTCTTCTCA
TCTACGGAGGATTTAATCTTATCTTATAAAACCTAAACGACGCTTGTCTTACTCCAAA
TTTCGACATTTGACGGCGAATCAAGCGATTCTAGAAGCGACGAGAGAAGTCGAACAAGATT
CACATCGTCGATTTTGAATCGTTCAAGGTATACAATGGCCTGCTCTTCTTCAAGCTCTA
GCTACTCGTACTTCTGGTAAACCCACTCAAATCCGGGTCTCGGGTATACCCGCTCCATCT
CTCGGTGAATCTCCGGAACCGTCGTTAATCGCCACCGGAAACCGCTCCGTGATTTCCGCC
AAGGTTCTGGATCTGAATTTGATTTTCATCCCAATTCTCACTCCCATACATTTACTTAAC
GGGTCAAGTTTCCGGGTGACCCCGGATGAAGTACTGGCCGTGAATTTTCATGCTCCAGCTC
TACAATACTCGACGACGCGGACGATAGTTGACACCGCACTACGGCTCGCCAAATCG
TTGAACCCGAGGGTCGTCACTCTCGGAGAATACGAAGTGAGCTTAAACCGGGTCGGTTTC
GCTAACCGGGTAAAGAACGCGCTTCAATTCTATTCCGCGGTTTTCGAATCCCTTGAACCG

AACTTGGGGCGTGATTCCGGAGGAGAGAGTGAGAGTTGAGCGAGAGTTGTTCCGCCGAGAGA
ATCTCGGGTTTGATTGGACCGGAGAAAACCGGAATTCATAGAGAAAGAATGGAAGAGAAA
GAGCAATGGCGGGTATTAATGGAGAATGCCGGTTTGAATCGGTTAAGCTGAGTAATTAC
GCAGTGAGCCAAGCGAAGATTCTATTGTGGAATTACAATTACAGCAATTTGTATCAATT
GTTGAATCTAAGCCTGGCTTCATCTCTTTGGCCTGGAACGATTACCTCTCCTCACTCTT
TCTTCTCGCGATAA

>G312 Amino Acid Sequence (domain in AA coordinates: 320-336)
MAYMCTDSGNLMAIAQQVIKQKQQEQQQQHHQDHIQIFGINPLSLNPWPNTSLGFGLSG
SAFPDPFQVTGGGDSNDPGFPFNNLDHHHATTTGGGFRLSDFGGGTGGGEFESDEWMETL
ISGGDSVADGPDCTDWHNDPDIYVYGPDPFDTPSRLSVQPSDLNRVIDTSSPLPPPTLW
PPSSPLSIPPLTHESPTKEDPETNDSEDDDFDLEPPLLKAIYDCARISDSDPNEASKTLL
QIRESVSELGDPТЕРВАFYFTEALSNRLSPNSPATSSSSSTEDLILSYKTLNDACPYSK
FAHLTANQAILEATEKSNKIHIVDFGIQVQIWPALLQALATRTSGKPTQIRVSGIPAPS
LGESPEPSLIATGNRLRDFAKVLDLNFDFIPILTPIHLLNGSSFRVDPDEVLA VNFMLQL
YKLLDETPTIVDTALRLAKSLNPRVVTLGEYEVSLNRVGFANRVKNALQFYSAVFESLEP
NLGRDSEERVVRVERELFGRRIISGLIGPEKGTIHRERMEEKEQWRVLMENAGFESVKLSNY
AVSQAKILLWNYNYSNLYSIVESKPGFISLAWNDLPLLTSSWR*

>G1444 (192..1001)
AATCCCCATATCCTTCGCAAGACCCCTTCTCTATATAAGGAAGTTCATTTTATTTT
GACACGCTGACAAGCTGACTCTAGCATATCTGGCACCGGCGACCACTCTTTGGTG
AAAGATCCCAAAAATCAAAATCGAAAGAGAGAATAAATCAAAGGAAGAATCTTTATCT
GCTTTCTCTCGATGAGGATCCGGAAACGACAAGTGCCTCTTCTTTATCGTCTCTATTAC
CAGTTCCTCTATCAGATCTCTACTTTAACCCTCACCGACGCGCCACCGCGAGATACTTTC
CGCGTGGTTATAAAGACGGCGGTGATGATTTTGGTTCTCTTCAGCTTTCGCTTCCGCCGC
CGTCGAGATTTCTGATCTCGCTTATTCAAAGAGATTGATAAAGAAGAAGGAGGAGGTCA
AGGCTTTGGATGATGATAATGGTGATGTAGACGTCAAGAGTCGTACTGATGCATCGGGCA
GCAAGAATGTTAATCCCCGAGGAGAATCCGTCTCTTCAATACAAGTTGTCGAGAAGAATG
AAAAGGTTGTGCTTTGAGGAAGAGAAGAGGCTTTATCAACTTTGAGGATTACGAAGATG
AGGAAGATGAAGAAGCTAGTGGCGGTGGAGGCCGTATTAATAAGGGAAAAAGAAAGCGA
AAAAGAGCGGTGGTGGGTAGAGGAAGGATCACGGTGCAGCCGTGTTAACGGTAGAGGAT
GGAGATGTTGTGACAAACGCTTGTGGTTATTCTCTTTGTGAGCATCATCTCGGTAAAG
GAAGGTTAAGGAGCATGAACAAGAGTGCTGGTGGTGGTGGCGCGGAGAAAAAGGCGGTGG
TGGTGAAGTGAAGAAGAAGAGAGTAAAGCTTGCATGGTAAAGGCACGTTCAATAAGTA
GTTTGCTTGGACAAACCAGCACTAGTGGTGGTACTAGTGGTGATGTTGATCAGGGTGAGA
TAAGTGCACCTGCTGATCAGTTCGCTGCATGTGATAAGTAGGTCTGTTGATCAGATTTG
CATGTATATGGATATGTGTATGTTATGTACATGATGATAATGGGCATAGCGCGGCCGCT
CTAGACAGGCCTGGAACCGGATCTCTAGCTAGAGCTTTCGTTAGTATCATCGGGTTTAG
ACAACGTT

>G1444 Amino Acid Sequence (domain in AA coordinates: 168-193)
MRIRKRQVPLPLSSLLPVPLSDLYFNRSPTATARYFRGGYKDGDDFGSLQLSLPPPSQI
SDRLIQRDLIKKKEEVKALDDNDGVDVKSRTDASGSKNVNPRGESVSSIQVVEKNEKV
SLRKRRGFINFEDYEDEEASGGGGRINKGKKKAKKSGGLEEGSRCSRNVNRRGWRC
QQTLVGYSLCEHHLGKGRVRS MNKSGGGRGGEKKAVVVEVKKRVKLGVMKARSISLLG
QTSTSGGTS GDVDQGEISAPADQFAACDK*

>G801 (27..746)
GATAGTGATAACGAAATCCTAATTCATGGCCGACAACGACGGAGCAGTGAGTAACGGCA
TCATAGTCGAGCAGACGTCAAACAAAGGACCTCTTAACGCCGTTAAGAAACCAACGCTTA
AAGATCGACACAGCAAGTTGACGGAAGAGGAAGAAGGATTTCGTATGCCAATCATTTGCG
CAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTCAAGTCCGATGGTCAAACCATAG
AGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCACTGGAACCTGGCACTACTC
CGGCGAGTTTCTCCACTGCTTCTCTCTCCACTTCTTCTCCGTTTACTCTCGGGAAACGTG
TCGTCAGAGCGGAGGAAGGAGAATCCGGCGCGGAGGAGGAGGGTTAACAGTGGGAC
ACACAATGGGGACTTCGTTAATGGGTGGTGGTGGTTCTGGTGGTTTTGGGCTGTTCCGG
CGAGGCCGATTTTCGGACAAGTCTGGAGCTTTGCAACCGGAGCTCCACCGGAAATGGTTT
TTGCGCAGCAGCAGCAACCAGCTACACTTCTCGTCCGCCACCAGCAGCAACAGCAAGCTT
CCGCCGCCGAGCAGCTGCAATGGGTGAGGCTTCAGCAGCTAGAGTTGGGAATTATCTTC
CGGGTCATCATCTCAATTTGCTTGCTTCTTGTCTGGTGAGCTAACGGGTCCGGTCCGGA

GGGAAGACGACCACGAACCACGTTGAGAAATGGTATTGCTTTTTGGTAATGTATAGAAA
AATTCCTATGTTTTATGTCATCGAAAGTGTTAGAAAGTACCTCTAATTTGCGGTTTTCTT
TTGCTCCTTTTTTACTTAATTTAAGCTTATTGCTTGTGGATTAGGGTTTTAGGGTTTAA
GAATATTTGGTCTCGTTAATTTGTTTCGGAGAGTGATAGAAAGAGAGAGATTGATTGA
TTGTTGTACCTAAAACGCTATAAAAGCTCTGTTTTTACTAGCGAAAAAA

>G801 Amino Acid Sequence (domain in AA coordinates: 32-93)
MADNDGAVSNGIIVEQTSNKGPLNAVKKPPSKDRHSKVDGRGRRIRMPIICAAARVFQLTR
ELGHKSDGQTI EWLLRQAEPSIIAATGTGTTPASFSTASLSTSSPFTLGKRVVRAEEGES
GGGGGGGLTVGHTMGTSLSLGGGGSGGFVAVPARPDFGQVVSFATGAPPEMVFAQQQQPAT
LFVRHQQQQASAAAAAAMGEASAARVGNYPGHHLNLLASLSGGANGSGRREDDHEPR*

>G1950 (42..764)
CTGAATTCGAACTTTGGAGAAGAAGAAGCTTTGATCAATCATGGAAATTGCAACCGATA
CAGCAAAGCAGATGAGAGACGAAGAGTTGTTCAAAGCAGCGGAATGGGGAGATTTCATCGT
TGTTTCATGTCTATTATCTGAAGAACAGCTCTCTAAATCTCTCAATTTAGAAACGAAGATG
GTGCTCTCTCTCTCTCGATGTCGCTTCTCTCTCGCCATTCTCAAATAGTGAAGTTGTTAT
CAAGTTCAGATGAAGCAAAGACTGTAATCAATAGCAAGGATGATGAAGGATGGGCTCCTT
TGCATTCCGCTGCTAGCATCGGTAATGCTGAGCTCGTTGAGGTGCTTTTGACCAGAGGTG
CTGATGTCAATGCCAAAAATAACGGTGGTCGCACTGCTCTTCACTATGCTGCTAGCAAAG
GCCGTTTGGAGATTGCTCAGCTTTTATTAAACACACGGTGCAAAGATTAAACATCACAGACA
AGGTTGGTTGCACTCCGCTTCACAGGGCAGCAAGCGTGGGAAAGTTAGAAGTTTGTGAAT
TTCTTATTGAAGAAGGAGCAGAGATCGATGCTACGGATAAAATGGGTCAAACCTGCACTCA
TGCAATTCAGTTATCTGCGATGACAAACAGGTTGCGTTCTCTGCTTATAAGACATGGTGCG
ATGTGGATGTAGAAGACAAGGAAGGCTACACTGTTCTAGGCCGAGCTACCAATGAATTCC
GACCTGCACTTATCGATGCTGCTAAGGCCATGCTTGAAGGATAAAATGACTCTGGATTAC
TTTAAACTTACTAATCTGAGAGTTGTTTAGTTACTTAAAGGATTTTTCTTTACTGTA
TCATGTTTGCAAATGTTTCTGCTTATCAATTCATGTTCTGT

>G1950 Amino Acid Sequence (domain in AA coordinates: 65-228)
MEIATDTAKQMRDEELFKAAEWGDSSLFMSLSEELSKSLNFRNEDGRSLHVAASFHGS
MEIVKLLSSSDEAKTVINSKDDGWAPLHSAASIGNAELVEVLLTRGADVNAKNNGGRTAL
HYAASKRLEIAQLLLTHGAKINITDKVGCTPLHRAASVGKLEVEFLIEEGAEIDATDK
MGQ TALMH SVICDDKQVAFLLIRHGADV DVEDKEGYTVLGRATNEFRPALIDAAKAMLEG
*

>G958 (55..1950)
CGTCGACATGTTTCATATTTGTTTCTAGCTAAGAAGTTTGTATAAGGCAGTGGACATGGCT
CCTGTTTCAATGCCCTCCAGGTTTCCGGTTTCATCCAACAGACGAAGAGCTTGTCATATAC
TACCTCAAGCGAAAGATTAATGGTCGGACTATTGAGTTAGAGATAATACCCGAGATTGAT
CTTTACAAATGCGAACCTTGGGATTTACCTGGGAAGTCCTTGCTGCCAAGTAAAGACCTA
GAATGGTTCTTTTTTCACTCCTCGAGACCGGAAATATCCAACGGATCAAGAACAACCGG
GCGACCAAAGCAGGTTACTGGAAAGCCACCGGAAAGATCGTAAAGTGACTTCACATTCA
CGGATGGTTGGAACAAAGAAAACATTAAGTTTATTACCGAGGAAGAGCGCCTCATGGCTCT
CGTACCGATTGGGTCTGCACGAGTACCGTCTTGAAGAACAAGATGTGACTCTAAATCC
GGTATACAGGATGCCATATGCACCTTGTTCGAGTATTTAAGAAGAGTGCTTTAGCCAACAAA
ATTGAAGAACCAACACCATGGTACGAAGAAGAACAAAGGAACGACTAATAGTGAACAATCT
ACTTCTAGTACTTGTTTGTATTCTGATGGAATGTATGAAAACCTCGAAAACCTCGGGGTAT
CCAGTCTCACCTGAGACAGGAGGCTTAACCTCAACTCGGTAATAATTCGTCGTCGGATATG
GAAACGATAGAGAATAAATGGAGTCAGTTTATGTCGCATGACACGTCCTTCAACTTCCCA
CCTCAGTCTCAATATGGAACAATCTCATATCCTCCTCGAAGGTTGATATAGCGTTAGAG
TGTGCAAGACTACAAATCGTATGTTGCCACCAGTACCACCACCTTACGTAGAAGGTCTC
ACACACAATGAATATTTTGGAAACAATGTAGCTAACGATACAGATGAAATGTTGAGCAAG
ATTATAGCATTTGGCTCAAGCCTCAGATGAGCCACGAACAGTCTAGACTCATGGGACGGT
GGTCTGCTTCCGGGAACCTTCCATGGAGACTTTAACTATTCCGGAGAAAAAGTCTCATGC
CTAGAGGCGAACGTGGAGGCTGTAGATATGCAAGAACACCATGTGAATTTTAAAGGAAGAA
AGACTTGTGAAAACCTTGAAGATGGGTAGGAGTATCAAGCAAGGAACCTTGAAGAGCTTC
GTTGAAGAACACTCAACGGTAATTCCTATAGAAGATATTTGGAGATATCATAATGATAAT
CAAGAACAAGAACATCATGATCAAGATGGTATGGACGTTAACAACAACAATGGAGATGTG
GATGATGCTTTCACTCGAGTTTTCGGAAAACGAACATAACGAGAATCTTTTGGACAAG
AACGATCATGAGACAACGAGTTCCTCATGTTTTGAGGTGGTAAAAAAGTTGAGGTTAGC

CATGGATTGTTTGT CACAAC TCGTCAGGTAACCAACACATTCTTCCAACAGATAGTACCA
TCGCAAACCGTTATAGTTTATATAAATCCGACGGATGGCAATGAGTGTTCATAGTATG
ACATCAAAAGAGGAGGTTTCATGTCCGTAAAAAGATAAATCCGCGAATCAACGGAGTAAGC
TCAACAGTTCTTGGACAATGGAGAAAATTCGCGCATGTTATTGGCTTCATTCTATGCTT
CTATTGATGCGTTGTGTTTCATCGAGGTAACTCTAACAAAAACAGAGGCAGTGAAGGTTAC
TCGAGGCAGCCTACGAGAGGAGATTGTAACAATCGGGGAACAATACTCATGATGGAAAAT
GCTGTCGTGAGAAGAAAAATTTGGAAGAAGAAGAGAAAAATATGGTTGACGAACAA
GGTTTTTCGTTTCAAGATAGTTTCGTATTGAAGAAGTTGGGGCTTTCTCTTGCTATCATC
TTAGCTGTTTCTACCATAAGTCTTATTGAATACTGAGGTTCAATATATCATATATGGCT
TTTCACTTTTCTATTGTACTCCCATTGCGCTAGGTCGTATGC

>G958 Amino Acid Sequence (conserved domain in AA coordinates: 7-156)

MAPVSMPPGFRFHPTDEELVIYYLKRKINGRTIELEIPEIDLYKCEPWLDPGKSLPSK
DLEWFFFSRDRKYPNGSRNRTKAGYWKATGKDRKVTSHSRMVGTKKTLVYYRGRAPH
GSRTDWMHEYRLEEQECDSKSGIQDAYALCRVFKKSALANKIEEQHHGTTKKNKGTNSE
QSTSSTCLYSDGMYENLENSGYPVSPETGGLTQLGNNSSSDMETIENKWSQFMSHDSFN
FPPQSQYGTISYPPSKVDIALECARLQNRMLPPVPPLYVEGLTHNEYFGNNVANDTDEML
SKIIIALAQASHEPRNSLSDWDGGSASGNFHDGDFNYSGEKVSLEANVEAVDMQEHVNFK
EERLVENLRWVGVSSEKELEKSFVEEHSTVPIEDIWRYHNDNQEQEHHDQDGMVNNNG
DVDDAFTLEFSENEHNENLLDKNDHETTTSSSCFEVVKKEVSHGLFVTTRQVTNTFFQOI
VPSQTVIVYINPTDGNECCHSMTSKEVHVHRKKINPRINGVSSSTVLGQWRKFAHVIGFIP
MLLLMRCVHRGNSNKNRSGEGYSRQPTRGDCNNRGTILMMENAVVRRKIWKKKKEKNMVD
EQGFRFQDSFVLKKLGLSLAILAVSTISLI*

>G1037 (1..1722)

ATGACTGTTGAACAAAATTTAGAAGCTTTGGATCAGTTTCCTGTAGGAATGAGAGTTCTT
GCTGTTGATGATGACCAAACCTTGTCTCAAAATCCTTGAATCTCTCCTTCGTCACTGCCAA
TACCATGTAACAACGACGACCAAGCACAAAAGGCTTTAGAGTTATTGAGAGAGAACAAG
AACAAGTTTGATCTGGTTATTAGTGATGTTGACATGCGCTGACATGGATGGTTTCAAATC
CTTGAGCTTGTGGTCTTGAAATGGACCTACCTGTCATAATGTTGTCTGCGCATAGTGAT
CCAAAGTATGTGATGAAGGGAGTTACTCATGGTGCTTGTGATTATCTACTGAAGCCGGTT
CGTATTGAGGAGTTGAAGAACATATGGCAACATGTCGTGAGAAGTAGATTTGATAAGAAC
CGTGGGAGTAATAATAATGGTGATAAGAGAGATGGATCAGGTAATGAAGGTGTTGGGAAT
TCTGATCCGAACAATGGGAAAGGTAATAGAAAACGTAAAGATCAGTATAATGAAGATGAG
GATGAGGATAGAGATGATAATGATGATTCTGTGCTCAAAAGAAGCAACGTGTTGTTTG
ACTGTTGAGCTGCATAAGAAATTTGTGAGCTGTTAACCCTTGGGATATGAGAAGGCT
ATGCCCTAAAAGATTTTGGATCTGATGAATGTTGAGAAGCTCACTAGAGAAAATGTGGCC
AGTCATCTTCAGAAAATCCGCCTTACTTGAAGAGGATCAGTGGTGTTGGCTAATCAGCAA
GCTATTATGGCAAACCTGAGTTACATTTTATGCAAATGAATGGACTTGATGGTTTCCAT
CACCGCCCAATCCCTGTTGGATCTGGTCAGTACCATGGTGGGGCTCCTGCAATGAGATCT
TTCCCTTCAAACCGGATTTCTTGGCAGACTCAATAGCTCTTCGGGGATCGGTGTCCGCAGC
CTTTCTTCTCCTCCTGCAAGGAATGTTCTTGCAAAACAGACCGATATCGGAAAGTTTAC
CATGTCTCATCACTTCTCTTAACCACAGTGATGGAGGAAACATACTTCAAGGGTTGCCA
ATGCCCTTAGAGTTGACACAGCTTCAGACAAACAACAACAAAGTAGAAACATGAACAGT
AACAAGAGCATTGCTGGGACCTCCATGGCTTTCTTAGCTTCTCTACGCAACAAAACCTCG
CTCATCAGTGCTCCTAATAACAATGTCGTGGTTCTAGAAGGTCACCCACAAGCAACTCCT
CCAGGCTTCCCAGGACACAGATCAATAAACGTTTGGAGCATTGGTCAAATGCTGTATCC
TCTTCGACTCACCCTCTCCCGGCACATAACAGTAATAGTATCAATCATCAGTTTCGAT
GTCTCTCCATTACCGCATTCTAGACCCGACCCCTTGAATGGAACAATGTGTATCAAGC
TACTCTATAACATTCTGTGACTCTGCCAATACATTGAGTTCTCCAGCCTTGATACAACA
AATCCCGAGCTTTCTGTAGAAACACGGACTTCGATTCAAACACAAATGTGCAACCTGGA
GTCTTTTATGGTCCATCCACGGATGCTATGGCTCTGTTGAGTAGTAGTAACCCGAAAGAA
GGGTTGCTGCTAGGCCAACAGAAGTTACAGAGTGGTGGATTATGTTGTCAGATGCTGGT
TCCTTAGATGATATAGTCAACTCCACGATGAAGCAGGTGTGA

>G1037 Amino Acid Sequence (domain in AA coordinates: 11-134, 200-248)

MTVEQNLEALDQFPVGMRLAVDDDDQTLKILESLLRHCCYHVTTTNAQKALELLRENK
NKFDLVISDVMPPMDMGFKLLELVGLEMDLPVIMLSAHSDPKYVMKGVTGACDYLLKPV
RIEELKNIWQHVVRSRFDKNGRSGNNNGDKRDGSGNEGVSNDPNNKGKNGRKRKDQYNEDE
DEDRDNDNDSQAQKQKRVVWTVELHKKFVAAVNQLGYEKAMPKKILDLMNVEKLTRENV

SHLQKFRLLYLKRISGVANQQAIMANSELHFMQMNGLDGFHHRPIPVGSGQYHGGAPAMRS
FPPNGILGRLNSSSGIGVRSLSPPAGMFLQNQTDIGKFHVSSLPLNHSDDGNNILQGLP
MPLEFDQLQTNNNKSRNMNSNKS IAGTSMAPFSFSTQNSLISAPNNNVVLEGHPQATP
PGFPGHQINKRLEHWSNAVSSSTHPPPAHNSNSINHQFDVSLPHSRPDPLEWNNVSSS
YSIPFCD SANTLSSPALDTTNPRAPCRNTDFDSNTNVQPGVFYGPSTDAMALLSSSNPKE
GFVVGGQKLQSGGFMVADAGSLDDIVNSTMKQV*

>G2065 (33..1124)

AACCACACAAAACAAAACAAAACATATTGATGGGGATGAAGAAGGTAAAGCTATCTT
TGATAGCTAATGAAAGATCAAGGAAAACATCCTTCATGAAGAGGAAAACGGGATATTCA
AGAACTCCACGAGTTGTCAACTCTATGTGGTGTCCAAGCTTGTGCTCTCATCTATAGTC
CATTCAACCGGTTCCAGAGTCATGGCCGTCAAGGAAGGTGCTAAAAGGTAGCTTCAA
AGTTTCTGGAGATCCCGCGGACAGCCGAACCAGGAAGATGATGGATCAAGAAACCCATC
TTATGGAGAGGATTACCAAAGCAAAAGAGCAACTAAAGAATTGGCTGCTGAGAACCGAG
AATTACAGGTTAGACGATTTATGTTTGAATTGTGTTGAAGGCAAAATGTCCCAGTATCGTT
ATGATGCAAAAAGACCTTCAAGATTGCTATCTTGTATGAATCTATATCTCGATCAGCTTA
ACGGAAGGATCGAGTCCATTAAAGAAAACGGTGAGTCGTTGTGCTTCCGCTCTCTCCTT
TTCCTACTAGAATTGGTGTGACGAAATTGGTGATGAGTCGTTTTCCGACTCTCCTATTC
ATTCTACAACCTAGGTTGTAGATACTCCTAATGCTACCAATCCTCATGTTCTTGCGGGCG
ATATGACTCCTTTCTTGATGCGGACGCAATGCGGTAACGTGCTCCAGTCGATTTTCTG
ATCATATTCAATATGAAAATATGAATATGAGTCAAAATCTGCATGAACCGTTTCAACACC
TTGTTCTCTACTAACGTTTGTGATTTTATCAAAATCAGAATATGAATCAGGTTCAATACC
AGGCTCCTAATAATCTGTTTAAATCAGATTCAACGAGAATTCTACAACATAAATTTGAATC
TGAATTTGAATCTGAATTCAAATCAGTATCTGAATCAACAACATCATTATGAATCCGA
TGGTGGAAACAATATGAATCATGTTGGAGGGCGTGAAGCATTCTTTCTGAGACAGAA
ACTACTACAACATACTCACTACCAGCCGTTGATCTTGCTTCCACCAGTTACATGCCTT
CAACCACCGATGTTTATGATCCTTACATCAACAACATCTCTAATCACAAAAGACGGAGA
TTTTCTAGTTTAA

>G2065 Amino Acid Sequence (domain in AA coordinates: TBD)

MGMKKVKLSLIANERSRKTSFMKRKNGIFKKLHBLSTLCGVQACALIYSPFIPVPESWPS
REGAKKVASKFLEMPRTARTRKMDQETHLMERITKAKEQLKNLAAENRELQVRRFMFDC
VEGKMSQYRYDAKDLQDLLSCMNLVLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
DESFSDSPIHSTTRVVDTPNATNPVLAGDMTPFLDADANAVTAPSRFSDHIQYENMMS
QNLHEFPQHLVPTNVCDFYQONQNMNVQYQAPNNLFNQIQREFYNINLNLNLNLSNQYL
NQQSFMNPMVEQHMNHVGGRESIPFVDRNYNYNQLPAVDLASTSYMPSTTDVYDPYIN
NNL*

>G2137 (77..1123)

GGGATTTGACTTTAGCACTTCAAATCCAAAGCTAAAAGACAAAAGAATAGAGGTTTCG
ATTTGCATCTCCATTAATGGGCATCGATCTTTCTTAAAGCTCGAGGCCGAGGAGAAAA
GAAAGAGATAGAAGGATCGAAACATAGCCGTGAGAACAAAGAAGACGAAGAATGATGC
TAGTGGTGATGAAGATGAACAAATGGTGAAAGAAGACGAAGATGATTCTTCTTTTAGG
TTTAAGAACCCGAGAAGAAGAAAACGAACGTGAAGAGCTCTTGACGCTACAGATCCAGAT
GGAAGTGTGAAGAAGAGAATACTAGGTTGAGGAAGCTTGTGAGCAGACTCTTGAAGA
TTATCGTCATCTTGAGATGAAATTCCTCGTTATCGATAAAACCAAGAAGATGGATCTTGA
AATGTTCTCTGGAGTACAAGGCAACGATGTGTGATATAACAAGTAAGGCTCGGAAAAG
AGGAGCTGAGAGATCTCCGTCAATGGAAAGAGAAATAGGGCTTTCACTTTCTTAGAGAA
AAAACAGAAACAAGAAGAGAGCAAGAAGCTGTTCACTCTCATCAACAAAGATACAATAG
TAGCAGCTTAGATATGAATATGCCACGTATCATTTTCATCTTCTCAAGGTAATAGAAAGGC
CAGGGTGTCCGTGAAGGCGAGATGTGAGACCGCAACATGAATGATGGATGCCAATGGAG
GAAGTACGGTCAGAAAACCGGAAAGGGAATCCATGTCTCGAGCTTATTACCGATGCAC
CGTGGCTCCAGGATGTCCCGTTAGAAAACAGGTGCAAGGTGTTTAGAAGACATGTCAAT
ACTGATAACAACCTACGAAGGAACACATAACCATCCACTTCCGGTCGGAGCAACAGCCAT
GGCTTCCACTGCCTCTACTTCTCCATTCTTGTACTCGATTCCAGTGACAACCTCTCTCA
TCCTTCTTATTACAAACTCCTCAAGCCATAGACTCTTCTTTGATTACATACCCACAAAA
TAGCAGCTACAACATCGAACCATAAGAAGCTTGAACCTTTGATGGTCCATCTAGAGGAGA
TCACGTTTTCATCTTCTCAAAACCGATTAAATGGATGATGTAGAGTTTCTATATCTCTA
TGCTTGTCTTTGGTCCCATTATTGTGTCATTATGGATTCTTGCCTTTCTTCTGTTCTC
GTTTCTAACATTTATGTTTCGTATA

>G2137 Amino Acid Sequence (conserved domain in AA coordinates:109-168)

MGIDLSLKLEAEKKKEIEGSKHSRENKEDEEHASGDEDEQMVKEDEDDSSSLGLRTRE
EENEREELLQLIQMESVKEENTRLRLKLEQTLQEDYRHEMKFPVIDKTKMDLEMFLGV
QKRCVDITSKARKRGAERSPSMEREIGLSLSLEKKQKQEESSKEAVQSHHQRYNSSSLDM
NMPRIISSSQGNRKARVSVRARCETATMNDGQCWRKYQKTAKNPCPRAYYRCTVAPGC
PVRKQVQRCLEMSILITTYEGTHNHPLPVGATAMASTASTSPFLLLDSSDNLSHPSYYQ
TPQAIDSSLITYPQNSSYNRTIRSLNFDGPSRGDHSVSSQNRLNWMW*

>G746 (1..1311)

ATGGGTGAGGAGTTAGCTGACACAATGAACCTGGATTTGAATCTTGGGCCTGGTCCTGAG
TCTGATCTCCAACCTGCACCAACGAGACTGTGAATTTGGCTGATTGGACTAATGACCCG
CCTGAGAGATCTTCTGAAGCTGTGACAAGGATCAGGACTCGGCATAGGACACGGTTCAGA
CAGCTTAATCTCCCGATCCCGGTTCTATCTGAAACCCATACCATGGCTATAGAGCTCAAC
CAGTTGATGGGAAATTCTGTAAATAGAGCTGCTATGCAGACTGGTGAGGGTAGTGAAAGA
GGCAATGAGGATTTGAAAATGTGTGAGAATGGCGATGGAGCCCTTGGGGACGGTGATTG
GATAAGAAAGCGGATGTCGAGAAAAGCAGTGGCAGCGACGGTAACCTTTTCGATGTGAAT
ATATGTTTGGATTTGTCGAAGGAGCCGGTTCTCACCTGTTGTGGTCATCTTTACTGTTGG
CCTTGCTGTACCAATGGTTACAAATTTCCGATGCAAAGGAATGTCCTGTTGTAAAGGA
GAGGTGACCTCCAAAACCGTGACACCGATCTATGGACGTGGAACCAAGAGAGAAATT
GAAGAGAGTTTAGATCTAAGGTCCCATGAGACCACACGCGAGACGCTATTGAGAGCTTG
AGGAATACAATTCAAAGGTGCGCTTTTACAATACCAATGGAAGAAATGATTAGACGTATA
CAGAATAGGTTTGACAGGGATTCAACCCAGTCCCTGATTTTAGTAACCGAGAGGCATCA
GAAAGAGTCAACGATCGAGCAATTCGATCCTTAACCGGTTGATGACATCTAGGGGAGTT
AGATCAGAGCAGAACCAGGCTAGTGCTGCAGCAGCAGCCATTGTGCGAGCATCAGAGGAT
ATTGATCTAAATCCAAACATTGCTCCTGATCTTGAAGGAGAAAGCAACACGAGATTCCAT
CCTCTCTTGATCAGGAGACAGTTACAGTCGCACCGAGTTGCAAGGATCTCGACTTTCCT
TCTCGGTTGAGTTTCACTGAGAGGCTTGTGGATGCGTATTTTAGGACTCATCCGTTGGGG
AGGAACCAAGAGCAAAACCATCATGCTCCTGTTGTGGTTGATGATAGAGACTCATTC
TCAAGCATTGCAGCTGTTATAAACTCTGAGAGTCAAGTGATACTGCAGTTGAGATCGAT
TCTATGGCTCTTTGACATCGTCTCGAGGAGAAGGAATGAGAATGGTTCGAGGGTTTCT
GATGTAGACAGTGCAGATTCTCGTCCGCTAGGAGAAGGAGATTTACTTGA

>G746 Amino Acid Sequence (domain in AA coordinates: 139-178)

MGEELADTMNLDLNLGPGPESDLQAPNETVNLADWTNDPPERSSSEAVTRIRTRHRTFR
QLNLPPIPVLSETHMTAIELNLQLMGNSVNRAMQTEGSEGRNEDLKMCENGDGLDGV
DKKADVEKSSGSDGNFFDCNICLDLSKEPVLTCGGHLYCWPCLYQWLQISDAKECPVCKG
EVTSTKVTPIYGRGNHKKREIEESLDTKVPMRPHARRIESLRNTIQRSPFTIPMEEMIRRI
QNRFRDSTPVPDFSNREASERVNDRANSILNRLMTSRGVRSEQNQASAAAAAIVASED
IDLNPNIAPDLEGENTRFHPLLIIRQLQSHRVARISTFTSALSSAERLVDAYFRTHPLG
RNHQEQNHHPVVDVDRDSFSSIAAVINSESQVDTAVEIDSMALSTSSRRRNENGSRV
DVSADSRPPRRRRFT*

>G2701 (46..837)

GTGTTTGTAGTTGAAACTTATTCTTCCCTTTTTTTGTTTGTAGGTATGGAGACTCTGCAT
CCATTCTCTCACCTACCTATCTCTGACCACCGGTTTCGTTGTTCAAGAGATGGTGAGCTTA
CACAGCTCGAGTAGCGGTAGCTGGACTAAAGAAGAGAACAGATGTTTCAACGAGCTCTT
GCGATATACGCTGAAGACTCGCCTGATCGCTGGTTTAAAGTTGCTTCCATGATCCCTGGA
AAGACTGTTTGTATGTTATGAAGCAATATAGTAAGCTTGAAGAAGACGTTTTTCGATATT
GAAGCAGGACGTGTTCCCATTCCTGGTTATCCTGCAGCTTCTTCTCCCTTGGGGTTTGAC
ACGGACATGTGTCGTAACGGCCTAGTGGAGCTAGAGGATCTGATCAAGATCGAAAGAAA
GGAGTCCCTTGGACAGAGGAAGAACACAGGAGATTCTTGTAGGCCCTTCTCAAGTACGGT
AAAGGAGATTGGAGAAACATATCGAGAACTTCGTGGTGTCAAAGACGCCAACGCAAGTG
GCGAGCCACGCCCCAAAAGTATTACCAGAGACAGCTCTCCGGAGCCAAGGACAAACGCAAG
CCAAGTATCCATGACATCAACCGGCAATCTTCTCAATGCCAATCTCAACCGTTCCCTT
TCCGATCATAGAGATATTCTCCCTGATTTAGGGTTTATCGATAAGGATGATACGGAGGAG
GGAGTAATAATTATGGGTCAGAATCTCTCTCAGAAAATCTGTTTTCTCCATCACCAACT
TCATTGCAAGCTGCGATTAACTTCGCCGAGAAAATGTCCTTCAGTGCCCGGAGCTTAAGGC
AACATAGAATCCCCAACTCAGCGGC

>G2701 Amino Acid Sequence (domain in AA coordinates: 33-81, 129-183)

METLHPFSLPISDHRFVVQEMVSLHSSSSGSWTKEENKMFERLAIYAEDSPDRWFKVA

SMIPGKTVFDMVKQYSKLEEDVFDIEAGRVPIPGYPAASSPLGFDTDMCRKRPSGARGSD
QDRKKGVPTWEEHRRFLLGLLKYGKGDWRNISRNFVVSKTPTQVASHAQKYYQRLSGA
KDKRRPSIHDIITGNLLNANLNRFSFSDHRDILPDLGFIDKDDTEEGVIFMGQNLSSLENLF
SPSPTSFEAAINFAGENVFSAGA*

>G1819 (1..639)

ATGGAAGAGAACACGGCAACAACAACCTACCTGCCGCAACCATCGTCTTCCCAACTG
CCGCCGCCACCATTTGATTATCAATCAATGCCGTTGCCGTCATATTCAGTCCGCTGCCG
TACTCACCGCAGATGCGGAATTATTGGATTGCGCAGATGGGAAACGCAACTGATGTTAAG
CATCATGCGTTTCCACTAACCCAGGATAAAGAAAATCATGAAGTCCAACCCGGAAGTGAAC
ATGGTCACTGCAGAGGCTCCGGTCTTATATCGAAGGCCGTGTGAGATGCTCATTCTTGAT
CTCACAATGCCATCGTGGCTTCATACCGTGGAGGGCGGTCCGCAAACTCTCAAGAGATCC
GATACGCTCACGAGATCCGATATCTCCGCCGCAACGACTCGTAGTTTCAAATTTACCTTC
CTTGGCGACGTTGTCCCAAGAGACCCTTCCGTCGTTACCGATGATCCCGTGCTACATCCG
GACGGTGAAGTACTTCTCCGGGAACGGTGATAGGATATCCGGTGTTTGATTGTAATGGT
GTGTACGCGTCACCGCCACAGATGCAGGAGTGGCCGGCGGTGCCTGGTGACGGAGAGGAG
GCAGCTGGGGAATTTGGAGGAAGCAGCGCGGTAATTGA

>G1819 Amino Acid Sequence (domain in AA coordinates: 46-188)

MEENNGNNHXLPPQSSSQLPPPLYQSMPLPSYSLPLPYSPQMRNYWIAQMGNATDVK
HHAFPLTRIKKIMKSNPEVNMVTAEPVLISKACEMILDLTMRSLHTVEGGRQTLKRS
DTLTRSDISAATRSFKFTFLGDVVPDPSPVTDVPLHPDGEVLPPGTIVIGYPVFDNCN
VYASPPQMGEWPAVPGDGEAAAGEIGSSSGN*

>G1227 (372..1451)

TCTTCCGTGTGTTAACAGAAGTCCCCACAATTGTCTGTCTTCGCTGCGAGACAAAACCTGC
CACAGCCAATAATGTTTCTCTGAGGGACCTTGCTTCTGTGTCAGAGACTCGCTCTCTCTC
CTCTTCTGTCTGTCTCAGCTCTCTCACCAACTCATCTTCAGTCTCAAACAAACATCTG
TTCTCATCTTTGTTTTCTTTCTTTCTTTCTCATATCTCATTTTCAATTTTCCCAATTTT
TCTTCAACATCTTCATAGCAATTTAAGACCACTATTCCATTATAAAGCTAACTGCTTTAG
AAACTCCTCACATTTATTTCTTCCCCATCATTGTTTATAGAGGGAGAAAGAAAAGAGC
TCAGCTTTCTGATGGAGAGGAGTATTCAAGGACAAAACAAGCTCTGTTGTTGGACAAA
AAGTGAATGTGAGAAGAAGCCTACAAGTTCAAGAACTGTAGAGGATCATCAAAGCTTTG
CCCTTGAAGAGGAAGAACAACAACCTCTCAACTCCGAGCTTGCTGCAAGACACAACAATAC
CATTTCTACAAATGCTGCAACAAGTGAAGACCCTTACCCTTTTTGTTCATTCAAAGACC
CAAGCTTTCTAGCACTACTATCTCTCCAGACACTTGAAAAGCCTTGGGAACTCGAAAAC
ACCTCCCACATGAAGTTCCAGAGTTTCATTACCGATCCATTCTGAAACCAACCACTACT
ATCATAATCCATCTTTGGGAAGGAGTCAATGAAGCCATCTCAAACCAAGAACTTCCATTCA
ACCCACTAGAGAATGCGCGTTCAAGACGCAAGCGGAAAAACAACAACCTTGGCATCATTGA
TGACAAGAGAAAAGCGAAAGAGAAGAAGAACTAAACCAACAAGAACATAGAAGAGATAG
AGAGTCAAAGAATGACACACATTGCGGTTGAACGAAACCGCAGACGCCAAATGAACGTTT
ATCTGAACCTCACTCCGCTCCATCATTCCATCTTCATACATCCAGAGGGGAGACCAAGCGT
CAATAGTAGGAGGAGCAATAGACTTCGTAAAGATCCTAGAGCAACAGTTGCAATCCCTTG
AAGCACAAAAGAGAAGTCAACAGAGTGATGATAACAAGAGCAAAATCCAGAAGATAACA
GTCTCAGGAACATTTTCGTGCAACAAGTTGCGTGCGAGTAATAAAGAAGAACAAAGTAGCA
AACTCAAATCGAAGCCACAGTGATAGAGAGTCAAGTCAACCTAAAAATTCATGTACGA
GGAAACAAGGACAACCTTCTCAGATCAATCATATTGCTGGAGAACTTCGATTCACTGTTT
TTCATCTCAACATCACATCTCCGACCAATACATCTGTCTCTTATTCTTCAACCTCAAGA
TGGAAGATGAATGTAATTTGGGATCAGCGGATGAGATAACGGCGGCGATTTCGTGAGATT
TCGACAGCTGATTGACTAATCCAAGTAAAAAGTAAAAATAAAAAAGAAACGTTTACTTTG
GTAACCTCGTTTTTCATGATTAAATCTTTATTTGGTCTGATGTGATTGGAGTCTTCTCGG
CATGGAACCTTGACTTTTGGTTTTAGGGTACTAGTCTCTACAGAAGCTGTGGTCTCTTTG
GATGC

>G1227 Amino Acid Sequence (domain in AA coordinates: 183-244)

MERSIQGNKLCCLDQKVNRRSLQVQETVEDHQSFALBEEBEEQQLSTPSLLQDTTIPFLQ
MLQQSEDPSPFLSFKDPSFLALLSLQTLKPEWELNYLPHEVPEFHSPHSETNHYHNP
SLEGVNEAISNQELPFNPLENARSRRKRKNNNLASLMTREKRKRRTKPTKNIEEIESQR
MTHIAVERNRNRQMNVLNLSLRSIIPSSYIQRGDQASIVGGAIDFVKILEQQQLQSLEAQK
RSQSDDNKEQIPEDNSLRNISSNKLRSNKKEQSSKLKIEATVIESHVNLIKIQCTRKQK
QLRSIILLEKLRFVTLHLNITSPTNTSVSYSFNLKMEDECNLGSADEITAAIRQIFDS*

>G2417 (118..1311)

CATACCGGTGGAAGATTCTGCTTTACTACGCTCTCCGCTTCTTCTCTCCTCGATTTCGAT
TCTCCTCATGGGTTTATCATGAATTTTAGGTTTTGAGTAATTCAGAACTCGAGTGATG
ATCCCGAATGATGATGATGATGCAAAATTCATGAAGAATTATCCGTTAAATGATGATGAT
GCAAAATTCATGAAGAATTATCCGTTAAATGATGATGATGCAAAATTCATGAGAAATTAT
CCGTTAAGGTCAATTCGACGGAGCTTTACACACTTGTTCATTGATACCACCTTCTTTA
CCAAACCTTCAGAAAGCAGCAGCAGACATGTCCTTCAATTCAGAACTCAATCAAATCATG
GCAAGGCCTTGTGATATGCTCCCTGCCAATGGTGGAGCTGTTGGTCATAACCTTTTGTG
GAACCAGGATTCAACTGCCCGGAGACAACAGATTGGATTCCCTCTCCACTCCCCATATT
TATTTTCTTCGGGTTCTCCCAATCTAATAATGGAGGATGGTGTCAATTGATGAGATTAC
AAACAAAGTGACTTGGCACTTTGGTATGACGACTTGATTACCACTGATGAAGATCCACTC
ATGTCTAGTATCTTGGCGGATCTTCTCTTGACACTAATTTCAACTCAGCTTCAAAGGTC
CAGCAACCAAGTATGCAATCGCAGATTCAACAACCCCAAGCTGTTCTGCAGCAGCCTTCT
TCTTGTGTGGAATTGCGCCCACTTGATAGGACAGTATCCTCAAACAGCAACAACAATAGC
AACAGTAATAATGCAGCAGCAGCAGCTAAGGGACGTATGCGTTGGACGCCTGAACTTCAT
GAGGTTTTTGTGACGCTGTTAACCAGCTCGGTGGCAGTAATGAAGCAACTCCTAAAGGT
GTCTTGAAGCATATGAAAGTCGAAGGTTTGAATTTTTTCATGTCAAAAGTCATTTGCAG
AAATATAGAACAGCTAAATATATATACCAGTACCATCAGAAGGTTGCGCGGAGGCAAGGTTG
ACACCGCTTGAGCAAATTACATCTGATGATACGAAACGTGGGATAGATATCACTGAGACT
CTGCGAATTCAGATGGAACATCAGAAGAACTGCATGAGCAGCTTGAGAGTCTAAGAACA
ATGCAACTTCGGATAGAAGAGCAAGGAAAGGCGCTGTTGATGATGATTGAGAAGCAAAAT
ATGGGTTTTCGGCGGACCAGAAACAAGGAGAGAAAACAGTGCGAAAACGCCTGAAAATGGT
TCAGAGGAGTCGGAATCCCCGCGGCCAAAGCGTCCGAGAAATGAAGAATGAAGGAAACCT
TTCTTCGATGGTAGATCATAAACTGTGGTTTTGGTGGAGTTGTAGAGTATGACTTATT
AGGAGTAGAGCTTTTCAGTCTTCTTCAGGC

>G2417 Amino Acid Sequence (domain in AA coordinates: 235-285)

MIPNDDDDANSMKNYPLNDDDDANSMKNYPLNDDDDANSMENYPLRSIPTELSHTCSLIPPS
LPNPSEAAADMSFNSELNQIMARPCDMLPANGGAVGHNPFLPEGFNCPETTDWIPSLPH
IYFPSGSPNLIMEDGVDEIHKQSDLEPLWYDDLITDEDPLMSSILGDLILLDTNFNSASK
VQQPSMQSQIQPPQAVLQQPSSCVLRPLDRTVSSNSNNNSNSNNAAAAKGRMRWTPPEL
HEVFVDVAVNLGGSNEATPKGVLEKMKVEGLTIFHVKSHLQKYRTAKYIPVPSGSPPEAR
LTPLEQITSDDTKRGIDITETLRIQMEHQKLLHEQLBRLTMQLRIBEQGKALLMMIEKQ
NMGFGGPEQGEKTSAKTPENGSESESPRPRPRNEE*

>G2116 (104..1117)

TTTCATCTCCATCATTATCTCCATTGACATTGTTCTCAATTGCGAATAATAATCATAATTA
TTCACACAACCAAAGCATTCTCTCAGATTCTCTTAAAAAATGGAGAAATCAGATCC
TCCACCAGTCCCAAAGCCCGCGCCACTATTATCCCTCTCTCCGATCCAATTCCTAATGC
CGATCCGATTCCATCTCTCTCTTCCACCGCGGATCTCGCTCCGACGATATGTCCATGTT
CATGTTTCATGGATCCCTCTCTCTCCCGCCGACCACTTCTCTCGACGACCTTCCCTCCGA
CGACGATCTCTTCTCTCTTTTCATCGATGTGATAGCCTCACCTCTAATCCCAATCCCTT
TCAAAATCCTTCCCTCTCTCTCAACTCCGTTTCCGGCGCTGCTAATCTCTCTCTCTCTC
TTCTCTCTGTCCTCGCCACCGTCACAGCAATTCGTTGACGCTGGATGCGCCATGTATGC
CGGTGATATCATGGACGCTAAGAAAGCTATGCCTCTGAAAACTCTCTGAGCTTTGGAA
CATCGATCCCAAACGCGCCAAAAGGATTCTAGCGAATCGACAATCTGCAGCTCGATCCAA
AGAGAGAAAAGCTCGATACATTCAAGAACTTGAGCGCAAAGTTCAATCTCTTCAAACCGA
AGCTACCACTCTCTCTGCTCAGCTTACTCTCTACCAGAGAGACACAAATGGACTAGCAAA
CGAAAACACAGAGCTGAAACTTAGGTTGCAAGCAATGGAACAACAAGCTCAGCTTCGTAA
TGCTTTAAACGAAGGTTGAGGAAAGAAGTTGAAAGGATGAAGATGGAGACAGGAGAAAT
CTCTGGTAATTCAGATTCTGTTTATGATGGGAATGCAGCAGATTCAATCTCTCTCAAC
TTTCATGGCTATTCACCATATCATGGCTCAATGAACCTCCATGATATGCAGATGCATTCT
TAGTTTCAATCTATGGAGATGTCGAATCTCAAAGCGTGTGCGACTTTCTACAGAACGG
CCGAATGCAAGGCTGGAGATTAGTAGCAATAGCTCAAGCTTAGTCAAATCTGAAGGACC
TTCTCTCTCTGCTAGTGAGAGTAGCTCTGCCTATTGACGACAAGATTATGATGAGGCTCA
TTTTTCTG

>G2116 Amino Acid Sequence (conserved domain in AA coordinates:150-210)

MEKSDPPPVPKPGATIIPSSDPIPNADPIPSSSFHRRSRSDMSMFMDPLSSAAPPSS
DDLPSDDDLFSSFDVDSLTSNPNPFQNPSSLSSNSVSGAANPPPPSSRPRHRHSNSVDA

GCAMYAGDIMDAKKAMPPEKLSELWNIDPKRAKRILANRQSAARSKERKARYIQELERKV
QSLQTEATTLAQLTLYQRDTNGLANENTELKLRLQAMEQQAQLRNALNEALRKEVERMK
METGEISGNSDSFDMGQQIQYSSSTFMAIPPYHGSMNLHDMQMHSSFNPMEMSNSQSVS
DFLQNGRMQGLEISSNSSSLVKSEGPSLSASESSAY*

>G647 (1..948)

ATGATGATCGGCGAAAAATAAAACCGGCCACATCCAACGATCCATATCCCTCAATGGGAT
CAAATCAACGATCCAACCGGCCACATCTCTTACCATTCTCTTCCGTCAACCTTAACAGC
GTAAACGACTACCCACACTCTCCGTCACCGTATCTCGACTCCTTCGCTTCTCTTCCGT
TACCTCCCGTCAAACGAGTTAACAACGATTTCAGACTCATCAAGTGGCGACGAGTCATCA
CCACTACCGACTCATTCTCTCCGACGAGTTTCGCATCTACGAGTTCAAAATCCGGCGA
TGCGCTCGAGGTCGATCTCATGATTGGACGGAGTGTCCGTTTCGCACATCCCGGAGAAAA
GCTCGACGACGTGATCCGAGAAAGTTTCTTACTCCGGCACCGCTTGTCTGAGTTTCGT
AAAGGAAGTTGTAGAAGAGGTGATTTCGTGTGAGTTCTCTCATGGAGTTTTCGAGTGTGG
CTCCATCCTTCTCGTTACCGTACTCAGCCGTGTAAAGACGGAAGTAGCTGCCGAGAGA
ATCTGTTTCTTCGCTCATACGACGAGCAGTTACGTGTATTACCTTGTTCGTTAGATCCA
GATCTTGGATTCTTCTCAGGATTAGTACTTCTCCGACTTCGATTCTTGTTCCTCTTCG
TTTTACCACCGTCGGAATCTCCGCCGCTTCTCCGAGTACCGGTGAACCTATTGCGTCG
ATGAGGAAAAATGCAATTGAACGGAGGTGGTTGTTCGTGGAGTTCTCCGATGAGATCTGCA
GTTAGGTTACCTTTTTCGTCGTCTCTGCGTCCGATTTCAGGCGGCAACGTGGCCGAGGATA
AGAGAGTTTGAGATCGAAGAAGCTCCGGCGATGGAATTTGTGGAATCTGGGAAAGAGCTG
AGAGCGGAGATGTATGCAAGACTCAGTAGAGAGAACTCACTCGGTTGA

>G647 Amino Acid Sequence (domain in aa coordinates: 77-192)

MMIGENKNRPHTIHIPQWDQINDPTATISSPFSSVNLNSVNDYPHSPSPYLDSEFASLFR
YLPSNELTNDSDSSSGDESSPLTDSFSSDEFRIYEFKIRRCARGRSHDWTECPFAHPGEK
ARRRDRPKRFHYSGTACPEFRKGSCRRGDSCEFSHGVEFCWLHPSRYRTQPKDGTSCRRR
ICFFAHTTEQLRVLPCSLDPDLGFFSGLATSPTSILVSPSPSPSPSPSPSTGELIAS
MRKMQLNGGGCSWSSPMRSARVLPFSSSLRPIQAATWPRIREFEIEAPAMEFVESGKEL
RAEMYARLSRENSLG*

>G974 (377..1162)

AAAAAAAAGTTGATATACTTTCTGGTTTTCTCCTTAACTTTTATTCTTTACAAATCCAT
CCCCCTTAGATCTGTTTATTTCCCGCTACTTTGATTCAATTTCTGTTAGTAATCTGTCTTT
CGTATAGAAGAAAACGATTTCTTGGTTTGTATTTCTTAAAGAGATCAATCTTTTTTTA
TTTTTGATCTTCTTGTGTTTTTTTCTTTGTAGAATTAATCGTTTGTGAGGGTATTTTT
TTAATTCCTCCTCTCAGAAATCTACACAGAGGTTTTTTATTTTATAAACCTCTTTTTTCG
ATTTTCTTGAAAACAAAAAATCCTTTCTTTACTTTTTTACAAGAACAAAGGAAAAAA
TTTCTTTTTATTAGAAATGACAACCTTCTATGGATTTTTACAGTAACAAAACGTTTCAACA
ATCTGATCCATTCCGTGGTGAATTAATGGAAGCGCTTTTACCTTTTATCAAAAGCCCTTC
CAACGATTCATCCGCGTTTGCCTTCTCTACCCGCTCCAATTTATACGGGTTCGGATCT
CCACTCATTTTCTCACCATCTTAGTCCTAAACCGGTCTCAATGAAACAAACCGGTACTTC
CGCGGCTAAACCGACGAAGCTATACAGAGGAGTGAGACAACGTCACTGGGGAAAAATGGGT
GGCTGAGATTGTTTTACCGAGGAATCGAATCGACTTTGGCTCGGAACATTCGACACGGC
GGAGGAAGCTGCTTTAGCTTATGACAAGGCGGCTATAAGCTCCGAGGAGATTTGCGCG
GCTTAATTTCCCTGATCTCCGTCATAACGACGAGTATCAACCTCTTCAATCATCAGTCGA
CGCTAAGCTTGAAGCTATTTGTCAAACTTAGCTGAGACGACGCAGAAACAGGTGAGATC
AACGAAGAAGTCTTCTCTCGGAAACGTTTCATCAACCGTCGCACTGAAACTACCGGAGGA
GGACTACTCTAGCGCCGGATCTTCCGCGCTGTTAACGGAGAGTTATGGATCTGGTGGATC
TTCTTCGCGGTTGTTCGGAGCTGACGTTTGGTGATACGGAGGAGGAGATTACGCCGCCGTG
GAACGAGAACGCGTGGGAGAAGTATCCGTCGTACGAGATCGATTGGGATTCGATTCTTCA
GGTTTCGAGTCTTGTAAATTAGATGTTGCCATAGGGGTATTTTAGGGACTTTAGAGCTCT
CTGCGATGGAGTTTTTGGTCATTGCAGAGATTTTATTATTATTAAGGGGGTTTGTATGT
TAATATCAATAAGTTTATCTACTTTGATGTTAATTAGTGTTAATCTCTGCGTCGGTCCA
AGCTGTTTTTTTTTGGCATGCTTCGACCGTGTGAGATTTCTTATGTAATTTTTGTAGTTC
CTTGATTTTCTTAGTTCAAGTTAAATTGGCACAAAAAATAAAAAAAAAA

>G974 Amino Acid Sequence (domain in AA coordinates: 81-140)

MTTSMDFYSNKTFFQSDPFGGELMEALLPFIKSPSNDSSAFAPSLPAPISYGSDDLHSFSH
HLSPKPVSMKGTGTSAAKPTKLYRGVQRHWGKWVAEIRLPNRTRLWLGTFTAEAAAL
AYDKAAYKLGRGDFARLNFDPDLRHNDYQPLQSSVDAKLEAICQNLAEATTQKQVRSTKSS

SRKRSSTVAVKLPEEDYSSAGSSPLLTESYSGSGSSSPLSELTFGDTEEEIQPPWENENAL
BKYPSEIDWDSILQCSSLVN*

>G1419 (27..692)

GAAGACTCCAACATAATTCATCATCTATGGCTTCTTCACATCAACAACAGCAAGAACAAAG
ACCAGTCAGCTTTAGATCTCATAACCCAACACCTTCTTACTGATTTCCCTTCCTTAGACA
CCTTTGCCTCCACCATCCACCCTGCACCACCTCAACTCTAAGCCAACGCAAAACACCTC
TTGCCACTATAGCAGTTCCTACTACTGCACCGGTGGTTCAAGAGAATGATCAAAGGCATT
ACAGAGGCGTCAGGAGAAGACCATGGGGTAAGTATGCGGCTGAGATCAGAGACCCAAACA
AGAAAGGTGTTCTGTCTGGTTAGGCACCTTTTGACACAGCCATGGAAGCTGCAAGAGGTT
ATGACAAGGCAGCTTTTAAACTACGAGGAAGCAAAGCTATTCTTAACTTCCCACTTGAAG
CAGGAAAGCATGAGGACTTGGGAGACAACAAGAAGACTATTCTTTAAAAGCAAAGAGGA
AGAGACAGGTGACGGAGGATGAAAGCCAGCTGATCAGCCGTAAAGCTGTTAAGAGGGAAG
AAGCTCAGGTTTCAAGCTGATGCTTGTCCATTAACGCCATCAAGTTGGAAGGGGTTTGGG
ACGGAGCAGACAGTAAAGACATGGGAATATTTCCGTGCCTCTGTTATCTCCTTGTCCAT
CTCTTGGACACTCTCAACTCGTAGTTACTTAAGCTTCAGAGGGTCAAACCTGGAAAAAATC
AACATTGGATTGTTTCAAAGCTTCTAGATTAGCTGATTGTAAAAAATGTTTACTATA
TTCATTCTTCTTAAATGCAATTCTTCTACCTTCC

>G1419 Amino Acid Sequence (domain in AA coordinates: 69-137)
MASSHQQQEQDQSDLDLITQHLLTDFPSLDTFASTIHHCTTSTLSQRKPPLATIAVPTT
APVVQENDQRHYRGVRRRPWGKYAAEIRDPNKKGVRVWLGTFTDAMEAARGYDKAAFCLR
GSKAILNFPLEAGKHEDLDGNKKTISLKAKRKRQVTEDESQILSRKAVKREBAQVQADAC
PLTPSSWKGFWDGADSKDMGIFSVPLLSPCPSLGHSQLVVT*

>G1634 (22..855)

TTATCTCGTAGCCTTTAAACGATGGAGACTCTGCATCCACTACTCTCGCACGTGCCAACT
TCTGACCACCGGTTGTAGTTCAAGAGATGATGTGCTTGCAAAGCTCGAGCTGGACTAAA
GAAGAGAACAAGAAGTTTGTAGCGAGCTCTTGCTGTCTACGCTGATGACACGCCTGATCGC
TGTTTCAAAGTTGCTGCTATGATCCCTGGAAAGACCATATCAGATGTCATGAGGCAATAC
TCTAAGCTTGAAGAAGACCTCTTCGATATCGAAGCAGGACTTGTCCCGATCCCGGGTTAC
CGTTCAGTTACTCCTTGTGGATTGATCAGGTTGTGAGTCCACGTGACTTTGATGCGTAT
CGTAAACTTCTTAATGGAGCCAGAGGATTGATCAAGACCGTAGGAAAGGAGTTCCATGG
ACGGAGGAAGAACAAGGAACTTTGTGGGATCAAAAACACCAACTCAGGTTGCAAGTCATGCC
CAAAAGTACTACCAAAGACAGCTTTCGGTGCGAAAGACAAACGACGGCCTAGCATTAC
GACATCACCACCGTCAATCTTCTCAATGCCAATCTTAGCCGTCCATCGTCTGATCACGGT
TGCTTAGTCTCAAAACAGGCCGAGCCGAAACTAGGGTTCACCGACAGGGATAATGCAGAG
GAGGGAGTTATGTTTCTTGGTCAGAATCTATCCTCGTCTTCTCTTCTACGATCCTGCC
ATTAAGTTTTCCGGAGCAAATGTTTACGGTGAAGGAGTTACTGTATCTCACAAGATCTT
GAAACGAGAAAAATGAAATTTTGAATTTTAACTATTGCAACGAAACCATAATTGC

>G1634 Amino Acid Sequence (domain in AA coordinates: 129-180)
METLHPLLSHVPTSDHRFVVQEMMCLQSSWTKEENKKFERALAVYADDPDRWFKVAAM
IPGKTISDVMRQYSKLEEDLFDIEAGLVPIPGYRSVTPCGFDQVVS PRDFDAYRKL PNGA
RGFDQDRRKGVPTWEEHRRFLLGLLKYKGDWNRNISRNFVGSKTPTQVASHAQKYYQRQ
LSGAKDKRRPSIHDITTVNLLNANLSRPSSDHGCLVSKQAEPKLGFTDRDNAEEGVMFLG
QNLSSVFSSYDPAIKFSGANVYEGGGYICISQDLETRK*

>G1637 (1..954)

ATGGTGAAGGAGACGGTGACGGTGGCGAAAACGTGCTCACACTGTGGCCATAATGGCCAT
AACGCACGGACTTGTCTCAACGGCGTTAATAAGGCAAGTGTTAACTGTTTCGGCGTTAAT
ATATCGTCTGATCCGATTAGGCCGCTGAGGTAACGGCGTTAAGGAAGAGTCTTAGTTTG
GGAAACCTTGATGCTCTTCTCGCTAACGATGAAAGTAACGGTAGCGGTGATCCTATCGCC
GCCGTTGATGATACCGGTTATCATTCCGATGGTCAGATTCAATCCAAGAAGGGTAAACT
GCTCATGAGAAGAAAAAGGGGAAGCCATGGACGGAAGAAGAACATCGTAATTTCTTAATC
GGTTTAAACAAACTCGGAAAAGGAGATTGGAGAGGCATTGCAAGAGTTTTCGTGTCGACA
AGAACACCAACACAAGTCGCAAGTCATGCTCAGAAAATATTTTATTAGGTTAAACGTTAAC
GACAAGAGAAAAAGACGTGCTAGTCTCTTTGACATCTCTCTCGAAGATCAGAAGGAGAAA
GAGAGGAATCTCAAGATGCTTCAACAAAGACTCCACCTAAACAACCAATAACCGGAATT
CAACAACCGGTAGTACAAGGTCATACTCAAACCGAGATTTTCAACAGGTTTCAGAATTTA
TCAATGGAGTATATGCCAATCTACCAACCCATACCACCTTACTACAACCTTTCACCTATT

ATGTACCATCCAAATTATCCAATGTACTATGCCAACCTCAAGTACCGGTTAGGTTTGTTCATCCTTCTGGTATACCTGTTCCAGACATATACCGATTGGTTGCCTCTGTCTCAACCGAGTGAAGCTTCTAATATGACAAATAAAGACGGTTTGGATCTTCATATCGGTTTGCCTCCA
CAAGCTACTGGAGCTTCTGACTTGACTGGTCATGGCGTTATTCATGTGAAATGA
>G1637 Amino Acid Sequence (domain in AA coordinates: 109-173)
MVKETVTVAKTCSHCHGNHARTCLNGVNKASVKLFGVNISSDPIRPPEVTALRKSLSL
GNLDALLANDESNGSGDPIAAVDDTGYHSDGQIHSKKGKTAHEKKKGKPWTEEEHNRFLI
GLNKLKGDWGRGIKSFVSTRPTQVASHAQYFIRLNVNDKRKRASLFDISLEDQKEK
ERN SQDASTKTPPKQIPITGIQQPVVQGHQTQTEISNRFQNL S MEYMPIYQPIPPYYNFPPI
MYHPNYPMYANPQVPVRVHPSGIPVPRHIPIGLPLSQPSEASNMTNKDGLDLHIGLPP
QATGASDLTGHGVIHVK*

>G1818 (601..1161)
TAACAAATCAAATAATTAGAGAAATAACCAAAATTTAACTTTTAGAGGGACTACAGGATT
TGTACTTTGTACATTCATATATTATTGTTATATATCGTTTCATACATTAATTTGAACCAA
TGTAATTAAGTAAATTCATTTAACATCATGAGCAAATCTTATTAATTAATCTCTTAA
AATTTTGAAGCAAATTAAGCTTTTACATTTTGAACAAATCATTTTAAACAAGATA
TTCAAACTAAGTTTGTACAGCAAATTTTAACTTTCAATTTTATAGAGAAAAAGGTAT
TTTTTTTTTGTTCATTTTATAAGACTATTATTGGTATATAATATACACTTTAAGTA
AAAACAAATCTCTTTCTTTTCTTCTTATAATACCAACCACAAGTCTGTCAGTCACACA
CATACAGTTAATAACATTAATATTCTTAACAAACTACTAAATAGGTTGAGATTCATATA
TGTAAGAGATCACTTCTTAATCTTATCCTACCATATCTTATATACGCTTAATTTTCTT
TATATATGCAAACCTCCACATAAAAAATATCTCAAACCCAAACACTTCAAACAAAAA
ATGGAGAACAAACAACAACCAACCAACAGCCACCGAAAGATAACGAGCAACTAAAGAGT
TTCTGGTCAAAGGGGATGGAAGGTGACTTGAATGTCAAGAATCACGAGTTCCTCATCTCT
CGTATCAAGAGGATAATGAAGTTTGATCCGGATGTGAGTATGATCGCTGCTGAGGCTCCA
AATCTCTTATCTAAGGCTTGTAAGTGTGTCATGGACCTCACGATGCGTTTATGGCTC
CATGCTCAAGAGAGCAACCGATCACGATACCGAAATCTGATGTTGATGCCGTAGTGTCT
CAAACCGTCATCTTTGATTTCTTGCGTGATGATGTCCCTAAGGACGAGGGAGAGCCGTT
GTCGCCGCTGCTGATCCTGTGGACGATGTTGCTGATCATGTGGCTGTGCCAGATCTTAAC
AATGAAGAACTGCCGCCGGGAACCGTGATAGGAATCCCGTTTGTACGTTTAGGAATA
CACGCCGCACACCCGAGATGCCGAGGCTTGGACCGAGGAGGATGCGACTGGGGCAAAT
GGAGGAAACGGTGGGAATTAATATTGGATTTGGGTTTGTAAACCGCTGTTGTGAGAACTT
GAATTTCTTTTGTAGTTCTGCTTATGTTTCAATGTTATGTTTGTAGTTGTGAATGTA
TTTCTGTTGTTTGTCCAAAAAAGAAATGATTTCTGTTGTTGTTCTTCAAATGA
ATCTAATGGTTTATGAATATTGGCTTTAGATTAAATTTATGCATACAAAAACACAAGGATT
ACGGATAAAAAAGTCTCAGTTTACCCATGGAAACATAATCTTCTAGTGATTCTTATGA
GAGTAGAAAAAGATCATATATTATAATCTATTTCATAAGAGATAGGGTACTGTAAACAAG
GATGTTTATTCGGCTATTCTTTTAAATCACTTTTACTTGTCAAGACTCTTTTGT
GTTTGCAGCTTTTGTAGATTACATCTAGAGGCAACAAGATCCAGAGATCTAGCAAAA
AAAACCTATTTTGAAACCTGAATCTATTTTAAAAATTTTCAACTCATTTTTCGTTCTTA
TTCTTTGTTTTCACCGAATTTGGCGCACAAACGATTTATTGAATTTTGTCTTTCAAG
>G1818 Amino Acid Sequence (domain in AA coordinates: 36-113)
MENNNNNHQPPKDNELKSFWSKMEGDLNVKNHEFPISRIKIMKFDPDVSMIAAEAP
NLLSKACEMFVMDLTMRSLHAQESNRLTIRKSDVDAVVSQTIVFDLRRDDVPKDEGEFV
VAAADPVDDVADHVAVPDLNNEELPPGTIVIGTPVCYGLGIHAPHPQMPGAWTEEDATGAN
GGNGGN*

>G1820 (1..609)
ATGGCTGAGAACAAEAACAACACGGCGACAACATGAACAACGACAACCACAGCAACCA
CCGTCGTACTCGCAGCTGCCGCCGATGGCATCATCAACCCCTCAGTTACGTAATTACTGG
ATTGAGCAGATGGAACCGTCTCGGATTTCAAAAACCGTCAGCTTCCATTGGCTCGAATT
AAGAAGATCATGAAGCTGATCCAGATGTGCACATGGTCTCCGAGAGGCTCCGATCATC
TTCGCAAAGGCTTGCGAAATGTTTCATCGTTGATCTCACGATGCGGTCTGCGCTCAAAGCC
GAGGAGAACAAACGCCACACGCTTCAGAAATCGGATATCTCAACGACGTGGCTAGCTCT
TTACACCTACGATTCTTGTAGTTGTTCCCTAAGGACGAGTCTATCGCCACCGCTGAT
CCTGGCTTTGTGGCTATGCCACATCTGACGGTGGAGGATACCGCAATATTATTATCCA
CCGGGAGTGGTGTGGAACCTCTATGGTTGGTAGTGGAATGTACGCGCCATCGCAGGCG
TGGCCAGCAGCGGCTGGTGACGGGGAGGATGATGCTGAGGATAATGGAGGAAACGGCGGC

>G1903 (1..1200)

>G1903 Amino Acid Sequence (domain in AA coordinates: 134-180)

>G371 (1..582)

>G371 Amino Acid Sequence (domain in aa coordinates: 21-74)

>G597 (255..1310)

256

TGCAATTTACAATGGCTCTCCACCAGCAGCAGCAACACAGTCAAGCTCAACCTCAGCAGT
CTCAGAACAGGCCATTGTCTATTCGGTGGAGACGACGGAAGTCTCTTTACAAGCAGCCGA
TGAGATCAGTATCACCACCGCAGCAGTACCAACCCAAGTCAAGCTGGTGAGAATTCTGTCT
TGAACATGAACCTGCCCCGAGGTGAGTCTGGAGGCATGACTGGAAGTGAAGTGAGCCAG
TGAAAAAGAGGAGAGGTAGACCGAGGAAATATGGGCCTGATAGTGGTGAAATGTCACTTG
GTTTGAATCCTGGAGCTCCTTCTTTCACTGTCTAGCCAACTAGTAGCGGCGGCATGGAG
GAGAGAAGAAGAGAGGAAGACCTCCTGGTTCTTCTAGCAAAAGGCTCAAGCTTCAAGCTT
TAGGCTCGACTGGAATCGGATTTACGCCTCATGTACTTACCGTGTGGCTGGAGAGGATG
TATCATCCAAGATAATGGCGTTAACTCATAATGGACCCCGTGTGTGTGTCTTGTCTG
CAAATGGAGCCATCTCCAATGTGACTCTCCGCCAGTCTGCCACATCCGGTGGAACTGTTA
CATATGAGGGGAGATTGAGATTCTGTCTTTATCGGGATCTTTCCATTTGCTGGAGAACA
ATGGTCAAAGAAGCAGGACGGGAGGTCTAAGCGTGTCAATTATCAAGTCCGGATGGTAATG
TCCTCGGTGGCAGTGTAGCTGGTCTTCTTATAGCAGCATCACCTGTTTCAAGTGTGTGTTG
GGAGTTTCTTACCAGACGGAGAAAAAGAACCAAAACAGCATGTGGGACAAATGGGACTGT
CGTCACCCGTATTACCGCGTGTGGCCCCAACGACAGGTGCTGATGACTCCAAGTAGCCAC
AATCTCGAGGCACAATGAGTGAGTCATCTTGTGGAGGAGACATGGAAGCCCTATTTCATC
AGAGCACTGGAGGACCTTACAATAACACCATTAACTATGCCCTGGAAGTAGCCAAGTGATC
TGTGTCCGCTTAAACCAACAACCTTCCCGTTATTAGAGTGATTTATTCTACATTTGGTT
TAGACTTTCTAGTTCTGATGGTTATTTCTACAGTTGGTTTAGACTTTCTAGTTCTGTTCA
GACAAAAGGAGTTTGATAAATTGACCGACCTATTTTGTGTGTTTGAGGTACTTTTCAAGAC
CATAGGTGTTTCAAAAATTAGAAATGTTCTGTTTAAAAAA
>G597 Amino Acid Sequence (domain in AA coordinates: 97-104,137-144)
MSGSETGLMAATRESMQFTMALHQQQHSQAQPQSQNRPLSFGGDDGTALYKQPMRSVS
PPQQYQPNASAGENSVLNMLPGGESGGMTGTGSEPVKKRRGRPRKYGPSGEMSLGLNPG
APSFTVSPSSGGDGGGKKRRRPPGSSSKRLKLQALGSTGIGFTPHVLTVLAGEDVSSKI
MALTHNGPRAVCVLSANGAISNVTLRQSATSGGTVTYEGRFEILSLSGSFHLLNNGQRS
RTGGLSVSLSSPDGNVLGGSVAGLLIAASPVQIVVGSFLPDGEKEPKQHVGMGLSSPVL
PRVAPTQVLMTPSSPQSRGTMSSESSCGGHHGSPHISTGGPYNNNTINMPWK*
>G1009 (28..1704)
AAAAAAAAAAAAACCTATTCCCAAAGATGAAGAACAATAACAACAAATCTTCTTCTTCT
TCTAGCTATGATTCTTCTTTGTCTCCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TCTTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
TCT
TCCACTTCTCCAGTCGAACGGCAAGATGGGTACCGGGAGTTTACCCAGCGATGCCACG
GCGGTTCTTCCGTATACCCCGCGGTCTTAACTTGAAGTCTTCTCGGCGGAGGAGCC
TCAACGACGACAACAAGCAATGCAACAAGTGCAATCTTCTTGGCGGCGTTGTCTTCTCT
TCCGACCTACAGCCACCGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AAGTCAATAGCCGCTAGCTTCTTAGGAACTACTCCGGTGGACACTCGTGGAGGTCTCT
AGCGTACATAAACAACACCGAATCTTCTAGCTGTCTCAGAGGCTTCGCCTACTCCGAAG
AAGAACGTAGAGAGTTTGGACAACGTACCTCGATTATATAGAGGAGTCAAGACATAGA
TGGACTGGAAGATACGAAGCTCATCTATGGGATAATAGTTGCCGAAGAGAAGGCCAAAGC
AGAAAAGGAAGACAAGTTTATTTAGGTGGTTATGATAAGGAAGATAAAGCACTAGAGCT
TACGACCTTGCAGCTCTTAAGTATTGGGGTCTTACAACCTACGACTAATTTCCCGATATCA
AATTACGAATCTGAACCTGAAGAAATGAAACACATGACTCGACAAGAGTTTCGTTGCTTCT
TTAAGACGGAAGAGCAGTGGATTCTCTAGGGGTGCCTCCATGTACAGAGGCGTCACTAGA
CATCATCAGCATGGTTCGATGGCAGGCACGAATTGGAAGAGTTGCAGGCAACAAAGACCTT
TATCTTGGCACATTTAGCACTCAAGAGGAAGCTGCAGAAGCTTATGATATAGCAGCGATC
AAATTCGCGGTCTAAATGCAGTCACCAATTTTCGACATCAGTCGATATGATGTCAAATCA
ATTGCTAGCTGTAATCTCCCTGTGGGTGGACTAATGCTTAAACCTTCTCCAGCAACCGCA
GCGGCTGACAAACCGTTGATCTTTCTCCATCCGACTCTCCATCTCTAACCACACCGTCC
CTCAGTTCAATGTGGCAACACCGGTCAATGACCATGGAGGAACTTTTTACCACACTGGT
ATACCAATCAAACAGACCCGGCTGATCATTATTGGTCCAACATCTTTGGATTCCAGGCA
AACCCGAAAGCAGAAATGCGACCATTAGCAAACCTTTGGGTCCGATCTTCAACCTTCT
CCTGGTTATGCTATAATGCCGGTAATGCAGGAAGGTGAAAACAACCTTTGGTGGTAGTTTT
GTTGGGTCTGATGGGTATAACAATCATTCGCTGCATCGAACCCGGTCTCAGCAATCCG
CTGTCTCGACAACATAATGAGTAACGGTAACGAAGGGTATGGTGGAAACATAAAGTGG
ATTAATAACAACATTTCAAGTCTTACCAAACCTGCAAAATCAAATCTCTCTGTTTTGCAC

ACACCGGTTTTTGGGTTGGAATGAGTATTCACATCTTAGTGAGAACTAAAATAAATATGT
AGGAAAAAATAAGGCTCTGTTTGAAGAAATCAGATATTTCTTCTTAGATTATTTAAGT
AGTTTAAAAAATATTTTTTAAAGTGTTCACCTTTACGTTTGTCTGCTGACCACGAATT
TTGCTGGATCTGACAGTACTAATCTTTGTTTAAATGACCTTATGGGTTCCCTTTTTTACTT
TCCAGAACTTTTATTTACTTTTTTCTTCATTTTCTTCATTTTTTTTGTGTGGGACAAT
ATGAATGATTGAAGATGGAACCTGCTTGCATGTGAATAAACGAAAATCAAACNATCTTCG
GTAACCTTAAAAA

>G1009 Amino Acid Sequence (domain in aa coordinates: 201-277, 303-371)

MKNNNNKSSSSSYDSSLSPSSSSSHQNWLSFSLSNNNNFNSSSNPNLTSSSTSDHHHP
HPSHLSLFQAFSTSPVERQDGSPPVSPDATAVLSVYPGGPKLENFLGGGASTTTTRPMQ
QVQSLGGVVFSSDLQPPPLHPPSAAEIYDSELKSIASFLGNYSGGHSSEVSVHKQQPNP
LAVSEASPTPKKNVESFGQRTSIYRGVTRHRWTGRYEHLWDNSCRREGQSRKGRQVYLG
GYDKEDKAARAYDLAALKYWGPTTTTNFPISNYESELEEMKHMTRQEFVASLRRKSSGFS
RGASMYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTQEEAAEAYDIAAIKFRGLNAV
NFDISRYDVKSIAASCNLPVGGMLPKPSPATAAADKTVDLSPSDSPSLTTPSLTFNVATPV
NDHGGTFYHTGIPIKPD PADHYWSNIFGFQANPKAEMRPLANFGSDLHNPSPGYAIMPVM
QEGENNFGGSFVGS DGYNNHSAASNVPVSAIPLSSTTTMSNGNEGYGGNINWINNNISSY
QTAKSNLSVLHTPVFGL*

>G170 (1..1107)

ATGGGGATGAAGAAGGTGAAGCTATCTTTGATAGCTAATGAAAGATCAAGGAAAACATCC
TTCATAAAGAGGAAAGACGGGATTTTAAAGAACTCCACGAGTTGTCAACTCTGTGTGGT
GTCCAAGCTTGTGCTCTCATCTACAGTCCATTCATACCGGTTCCAGAGTCATGGCCGTCA
AGGGAAGGTGCTAAAAAGGTGGCTTCAAGGTTTCTGGAGATGCCGCCGACAGCCCGAACC
AAGAAGATGATGAGTCAAGAGACTTACCTTATGGAGAGGATTACCAAAGCAAAGAGCAA
CTAAAGAACCCTGGCTGCTGAGAACCGAGAGTTACAGGTTAGACGATTATGTTTGTATTGT
GTTGAAGGC AAAATGTCCCAGTATCATTATGATGCAAAAGACCTTCAAGATTTGCAATCT
TGTATAAATCTATATCTCGATCAGCTTAACGGAAGGATCGAGTCCATTAAAGAAAATGGT
GAGTCGTTGTTGTCTTCCGTCTCTCCTTTTCTTACTAGAAATTGGTGTGACGAAATTGGT
GATGAGTCATTTTCCGACTCTCCTATTATGCTACAACCTGGGTTGTAGATACTCTTAAT
GCTACCAATCCTCATGTTCTTACGGGCGATATGACTCCTTTTCTTGATGCGGACGCAACT
GCGGTAACCTGCTTCCAGTAGATTTTGTGATCATATTCATATGAAAATATGAATATGAGT
CAAAATCTGCATGAACCGTTTCAACACCTTGTTCCTACTAACGTTTGTGATTTTTTTCAA
AATCAGAATATGAATCAGGTTCAATACCAGGCTCCTAATAATCTGTTTAAATCAGATTCAA
CGAGAATTCTACAACATAAATTGAATCTGAATTTGAATCTGAATTCGAATCAGTATCTG
AATCAACAACAATCATTATCATGAATCCGATGGTGGAAACAACATATGAATCATGTTGGAGGG
CGTGAAAGCATTCCTTTCGTGGACGGAACTGCTACAACCTACCATCAACTACCATCCAAT
CAACTACGACCGTTGATCATGCTTCCACCAGTTACATGCCTTCCACCACCGGTGTCTAT
GATCCTTACATCAACAATAATCTCTAA

>G170 Amino Acid Sequence (domain in aa coordinates: 2-57)

MGMKKVKLSLIANERSRKTSPIKRKDGIFKKLHELSTLCGVQACALIYSPFIPVPESWPS
REGAKKVASRFLEMPPTARTKKMMDQETYLMEITKAKEQLKNLAAENRELQVRRFMFDC
VEGKMSQYHYDAKDLQDLQSCINLYLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
DESFSDSPIHATTGVVDTLNATNPHVLTGDMTPFLDADATAVTASSRFPDHIPIYENMNMS
QNLHEPFGHLVPTNVCDFFQONQNMNQVQYQAPNNLFNQIQREFYNINLNLNLNSNQYL
NQQSFNMNPMVQHMNHVGGRESIPFVDGNCYNYHQLPSNQLPAVDHASTSYMPSTTG
VYDPYINNLL*

>G1768 (185..1426)

CTTCCTTTTGCTTCAGCTGCGAGCTTTGGTTGGATCTCTCACTTGCAAAACCAAATCCCT
TATCGACTTCCACCGAAAGATCACTTCTTAACCTACACAAGGTGTTTGTATGAAGATCA
GATAAATAAAGGTCAATTGAGGATAATGGTTGATGTTCAAAGATTCTTACTTGCTTATT
TGTGATGGACAATGTAAGAGGTTCAATAATGTTGCAGCCACTGCCAGAGATAGCTGAGAG
TATCGATGATGCTATCTGCCATGAACCTCTCCATGTGGCCTGATGATGCTAAAGATTGTT
ATTGATAGTGGAGGCAATATCAAGGGGAGACTTGAAGTTGGTACTTGTGCTTGTGCAAA
AGCTGTTTCTGAGAATAATCTTCTAATGGCACGATGGTGTATGGGTGAGTTGCGCGGTAT
GGTTTTCGATTTCTGGTGAGCCAATCCAGAGATTGGGAGCTTATATGTTAGAAGGGCTGT
TGCTAGGCTTGCTGCTTCTGGTAGTTCGATATATAAGTCTCTCCAGTCCAGAGAACCAGA
GAGTTATGAATTTTATCTTATGTGTATGTTCTGCATGAGGTTTGTCCATATTTCAAGTT

TGGATACATGTCAGCGAATGGTGGCATTGTCAGAAGCAATGAAGGATGAAGAGAGGATTCA
CATTATTGACTTCCAAATTTGGACAAGGGAGCCAGTGGATAGCACCTATCCAGGCTTTTGC
AGCTAGGCCTGGTGGGGCTCCAAATATTGCAATTACCGGAGTTGGTGATGGATCTGTCTT
GGTTACAGTCAAGAAGAGACTAGAGAACTTGCAAAGAAGTTTGATGTTCCATTACAGGTT
CAATGCGGTTTCAAGGCCAAGTTGTGAAGTTGAAGTGGAAATCTTGATGTCCGAGATGG
CGAAGCCCTTGGAGTGAACCTTTGCTTACATGCTGCATCATTGTCAGATGAGAGTGAAG
CATGGAAAACACAGGGACCGGTTGCTGAGGATGGTGAAGAGTCTATCACCTAAAGTAGT
CACTCTTGTTGGAACAAGAATGCAACACGAACACTTCCCTTTCTCTCTAGGTTCTTGA
GACATTAAGTTATTACACGGCAATGTTTGAATCTATCGATGTTATGCTTCCGAGAAATCA
CAAGGAAAGGATCAATATCGAGCAGCACTGCATGGCAAGGGATGTCGTCAACATCATAGC
TTGTGAAGGAGCCGAGAGGATCGAAAGACACGAGCTTCTCGGGAATGGAAGTCAAGGTT
TTCCATGGCGGGTTTTGAGCCATACCCCTTGAGCTCAATCATTTCAGCCACCATTAGAGC
CCTCTTGAGAGATTACGCAACGGGTATGCGATTGAAGAAGAGATGGTGCTCTGTACCT
TGGTTGGATGGACCGAATCTTGGTCTCATCTTGTGCATGGAAGTGAAGAATAAACGCTCT
CCAAGAATGTAATGCAAAAGACAGAACTGGAAGTAATAGATAGTTTGTCTCATAACCAT
TAATAAGGTTGAATCAAATCATATACATCCCATGCTACAACATTATACACAGGCTCCATC
AACAAAGAGGGCTCTTGTGTGTACCTTCTCTCTGTAACCTTATTTGAACCAAAT
GGAAGTGGTTACAT

>G1768 Amino Acid Sequence (domain in AA coordinates: 54-413)
MDNVRGSI MLQPLPEIAESIDDAICHEL SMWPDDAKDLLLIVEAISRGDLKLVLVACAKA
VSENNLLMARWCMGELRGMVSI SGEPIQRLGAYMLEGLVARLAASGSSIIYKSLQSRPES
YEFLSYVYVLHEVCYPYFKFGYMSANGAIAEAMKDEERIHIIDFQIGQGSQWIALIQAF
RPGGAPNIRITGVGDGSLVLT VTKRLEKLAKKFDV PFRFNAVSRPSC EVEVENLDVRDGE
ALGVNFAYMLHLLPDESVMENHRDRLRMVKSLSPKVVT LVEQECNTNTSPFLPRFLET
LSYYTAMFESIDVMLPRNHKERINIEQHCMARDVVNIIACEGAERIERHEL LGKWSRFS
MAGFEPYPLSSII SATIRALLRDYSNGYAI EERD GALYLGMMDRILVSSCAWK*

>G185 (77..988)

ATGCAAAAATAAACATAGTAACAATACTTTAAACTATTTACACCACCTTTAATCTTATTCT
CCACTCTTTGAACGTAATGGAGAAGAACCATAGTAGTGGAGAGTGGGAGAAGATGAAGAA
CGAGATCAACGAGCTAATGATAGAAGGAAGAGACTATGCACACCAGTTTGGATCAGCTTC
ATCTCAAGAAACACGTGAACATTTAGCCAAAAAGATTCTTCAATCTTACCACAAGTCTCT
CACCATCATGAACACTCTCCGGCGAAGTTGACCAAGTTTCTCAGGGTGGAGGAAGCCCCAA
GAGCGATGATTCCGATCAAGAACCACTTGTCTCATCAAGAGTTCGAAGAAGTCAATGCCAAG
GTGGAGTTCAAAAGTCAGAAATTTGCCCTGGAGCTGGTGTGATAGAACGCTGGACGATGG
ATTGAGTTGGAGAAAGTACGGCCAGAAGGATATTCTCGGAGCCAAATTTCCAAGAGGATA
CTATAGATGCACGTATAGAAAGTCTCAAGGATGTGAAGCCACTAAACAAGTCCAAAGATC
TGATGAAAATCAGATGCTCCTTGAGATCAGTTACCGAGGAATACATTCTTGCTCTCAAGC
TGCAAATGTCGGTACAACATGCCGATACAAAACCTCGAACCAGAACAGACCCAAAGACA
CGGAAATCTTGACATGGTAAAGGAAAGTGTAGACAACACTACAATCACCAGCACATTTGCA
TCACAACCTTCACTATCCATTGTCTATCTACCCCAATCTAGAGAATAACAATGCCTATAT
GCTTCAAAATGCGAGATCAAAACATCGAATATTTTGGATCTACGAGCTTCTCTAGTGATCT
AGGAACTAGTATCAACTACAATTTTCCAGCATCTGGCTCGGCTTCTCACTCAGCATCAAA
CTCTCCGTCCACCGTCCCTTTGGAATCCCCGTTTGAAAGCTATGATCCAAATCATCCATA
TGGAGGATTTGGTGGGTTCTATTCTTAGTTATCTACTTAAGGGAGGGACGGAACTTTTTA
CATGACCTCTTGATTAAAGAGAGAGTTTTCATAATAGCTAATCAATTTCTTATTCAAATA
TCCGAGTTTTTTTTCTAATCATGTTTATCAATTGTCTTATTACAGAAGGCTTATTTTCAG
GTCTATGTTGAAATAAATGGATTTGTACTCGTAGGTATGATCCTTGTATCTAAAAA
AAAAA

>G185 Amino Acid Sequence (domain in AA coordinates: 113-172)
MEKNHSSGEWEKMKNEINELMIEGRDYAHQFGSASSQETREHLAKKILQSYHKS LTIMNY
SGELDQVSQGGGSPKSDSDQEPLVIKSSKKSMPRWSSKVRIAPGAGVDR TLDDGF SWRK
YGQKDILGAKFPRGYRCTYRK SQGCEATKQVQRSDENQMLLEISYRG IHS CSQAANVGT
TMPIONLEPNQTQEHGNLDMVKESVDNYNHQAHLHNLHYPLSSTPNLENNNAYMLQMRD
QNIEYFGSTSFSSDLGTSINYNFPASGSASHSASNSPSTVPLES PFESYDPNHPYGGFGG
FYS*

>G1931 (5..592)

ATCAATGGAAGGGGTTGACAACACAAATCCTATGTTAACCCTAGAAGAAGGCGAAAAACAA

CAATCCTTTTCTTCTTAGATGACAAAACATTAATGATGATGGCTCCTTCGTTAATCTT
TTCGGGCGATGTAGGTCCATCTTCTTCTTCTTGTACTCCAGCAGGTTATCATCTATCTGC
TCAGCTGGAGAACTTTCGAGGAGGTGGAGGAGAGATGGGAGGATTAGTGAGTAATAATAG
CAATAATAGTGATCATAATAAGAATTGCAACAAAGGAAAAGGGAAGAGAACTTTGGCAAT
GCAGAGGATAGCTTTTCATACAAGGAGTGATGATGATGTTCTTGATGATGGTTATCGTTG
GCGAAAGTACGGTCAGAAATCTGTCAAGAACAATGCTCATCCCAGGAGCTATTATAGATG
TACATACCACATGCAACGTGAAGAAACAAGTGCAAGACTGGCAAAAGATCCAAACGT
TGTCGTAACAACCTACGAAGGTGTTTCATAATCATCCTTGTGAGAAGCTCATGGAGACTCT
TAGCCCTCTCCTTAGGCAACTTCAGTTCCTCTCAAGAGTTTCTGATCTGTAATTATTGAA
TGTTAATTAGTGGTGTAATACATTAATTATGCTTTAATCTCTCCATTGACCCCTCAATC
>G1931 Amino Acid Sequence (domain in AA coordinates: 114-170)
MEGVDNTNPMLEEGENNPFSSLDKTLMMAPSLIFSGDVGPSSSSCTPAGYHLSAQ
LENFRGGGEMGGLVSNNSNNDHNKNCNKKGKRTLAMQRIAFHTRSDDDVLDGYSR
KYGQKSVKNNAHPRSYRCTYHTCNVKKQVORLAKDPNVVVTYEGVHNHPCEKLMETLS
PLLRQLQFLSRVSDL*

>G2543 (1..2169)
ATGAGTTTCGTCGTCGGCGTGGGCGGAAGTGGTAGTGGGAAGCGGCGGAGACGGTGGTGGT
AGTCATCATCAGCAGCGCTTGAACTGATAGGAAGAAGAAACGTTACCATCGTCACACC
GCTCAACAGATTCAACGCTTGAATCGAGTTTCAAGGAGTGCTCTCATCCAGATGAGAAA
CAGAGGAACAGCTTAGCAGAGAATTGGGTTTGGCTCCAAGACAAATCAAGTTCTGGTTT
CAGAACAGAGAAGCTCAGCTTAAGGCTCAACATGAGAGAGCAGATAATAGTGCACTAAAG
GCAGAGAATGATAAAATTCGTTGCGAAAACATTGCTATTAGAGAAGCTCTCAAGCATGCT
ATATGTCCTAACTGTGGAGGTCTCTCTGTTAGTGAAGATCCTTACTTTGATGAACAAAAG
CTTCGGATTGAAAATGCACACCTTAGAGAAGAGCTTGAAAGAATGTCTACCATTGCATCA
AAGTACATGGGAAGACCGATATCGCAACTCTCTACGCTACATCCAATGCACATCTCACCG
TTGGATTGTCAATGACTAGTTTAACTGGTTGTGGACCTTTTGGTCATGGTCTCTCACTC
GATTTTGTATCTTCTCCAGGAAGTTCTATGGCTGTTGGTCTTAATAATAATCTGCAATCT
CAGCCTAACTTGGCTATATCAGACATGGATAAGCCTATTATGACCGGCATTGCTTTGACT
GCAATGGAAGAATTGCTCAGGCTTCTTCAGACAAATGAACCTCTATGGACAAGAACAGAT
GGCTGCAGAGACATTCTCAATCTTGGTAGCTATGAGAATGTTTTCCCAAGATCAAGTAAC
CGAGGGAAGAACCAGAACTTCGAGTCGAAGCATCAAGGTCTTCTGGTATTGTCCTCATG
AATGCTATGCACTTGTGCGATGTTTCATGGATTGTGTCAAGTGGACAGAACTCTTCCC
TCTATCATTCAGCTTCTAAAACACTTGCAGTGATTCTTTCAGGAATGGGAGGTACCCAT
GAGGGTGCATTGCATTTGTTGTATGAAGAAATGGAAGTGCTTTCGCCTTTAGTAGCAACA
CGCGAATTCTGCGAGCTACGCTATTGTCAACAGACTGAACAAGGAAGCTGGATAGTTGTA
AACGCTCTCATATGATCTTCTCAGTTTGTCTTCTCACTCTCAGTCTTATAGATTTCCATCT
GGATGCTTGATTAGGATATGCCAATGGATATTCGAAGGTTACTTGGGTTGAACATATT
GAACTGAAGAAAAGAACTGTTTCATGAGCTATACAGAGAGATTATTACAGAGGGATT
GCTTTTGGGGCTGATCGTTGGGTTACCACTCTCCAGAGAATGTGTGAAAGATTGCTTCT
CTATCGGTACCAGCGTCTTCATCTCGTGATCTCGGTGGAGTGATTCTATCACCGGAAGGG
AAGAGAAGCATGATGAGACTTGCTCAGAGGATGATCAGCAACTACTGTTTAAAGTGTCAGC
AGATCCAACAACACACGCTCAACCGTTGTTTCGGAAGTGAACGAAGTTGGAATCCGTGTG
ACTGCACATAAGAGCCCTGAACCAACCGGCACAGTCTATGTGCAGCCACCACCTTCTGG
CTTCCCAATTCTCTCAAAATGTCTTCAATTTCTCAAGACGAAAGAACCCTGCTCAG
TGGGATGTTCTTTTCAAACGGAACGAGTGCAAGAAGTTGCTCACATCTCAAACGGATCA
CATCTTGGAACTGCATATCGGTTCTACGTGGATCCAATGCAACACATAGCAACAACATG
CTTATTCTGCAAGAAAGCTCAACAGACTCATCAGGAGCAATTTGTGGTCTACAGTCCAGTG
GATTTAGCAGATTGAACATCGCAATGAGCGGTGAAGATCCTTCTTATATTCTCTCTTG
TCCTCAGGTTTCAATCTCACCAGATGGAAATGGCTCAAACCTGAAACAAGGAGGAGCC
TCGACGAGCTCAGGACGGGCATCAGCTAGCGGTTGCTTGATAACGGTTGGGTTTCAGATA
ATGGTAAGCAATTTACCGACGGCAAACTGAATATGGAGTGGTGGAAACGGTTAATAAC
CTGATAGGAACAACCTGTACATCAAATTAACCGCCTTGAGCGGTCTACAGCTTCAACT
ACAGCTTGA

>G2543 Amino Acid Sequence (domain in AA coordinates: 31-91)
MSFVVGVGSGSGSGDGGGSHHHDGSETDRKKRYHRHTAQQIQRLESSFKECPHPDEK
QRNQLSRELGLAPRIKFWFQNRRTQLKAQHERADNSALKAEENDKIRCENIAIREALKHA
ICPNCGGPPVSEDYPFDEQKLRIENAHLEELERMSTIASKYMGRPISQLSTLHPMHISP

LDLSMTSLTGCGPFGHGPSLDFDLLPGSSMAVGPNNNLQSQPNLAISDMDKPIMTGIALT
AMEELLRLQLTNEPLWTRTDGCRDILNLGSYENVFPRSSNRGKNQNFVEASRSSGIVFM
NAMALVDMFMDVCWKTELPFSIIAASKTLAVISSGMGGTHEGALHLLYEEMEVLSPLVAT
REFCELRYCQQTEQGSWIVNVSYDLPQFVSHSQSYRFPSPGCLIQDMPNGYSKVTWVEHI
ETEEKELVHELRYEIIHRGIAFGADRWVTTLQRMCFASLSVPASSSRDLGGVILSPEG
KRSMMLRAQRMISNYCLSVRSNNRSTVSELNEVGIRVTAHKSPEPNGTVLCAATTFW
LPNSPQNVFNFLKDERTRPQWDVLSNGNAVQEVAHISNGSHPGNCISVLRGSNATHSNNM
LILQESSTDSSGAFVVYSPVDLAALNIAMSGEDPSYIPLLSSGFTISP DGNGSNSEQGGA
STSSGRASASGLITVGFIQIMVSNLPTAKLNMESVETVNNLIGTTVHQIKTALSGPTAST
TA*

>G264 (30..1430)

CTTGTACAGTTTCTGATTAGATTCAACAATGAACGGCGCATTAGGTAACTCCTCCGCCT
CCGTTAGCGGCGGAGAAGGAGCCGAGGACCAGCGCCTTTCTTGGTGAAAACCTACGAGA
TGGTCGACGATTTCATCAACGGACAGATCGTATCGTGGAGCGCTAACAAACAGCTTCA
TCGTTTGGAAATCATGCCGAATTTTCACGCCTCCTTCTTCCAACCTACTTCAAACACAATA
ACTTCTCTTCCTTCATTTCGTAGCTCAATACCTATGGGTTTAGGAAGATTGATCCAGAGA
GGTGGGAGTTTTTGAATGATGATTTTATTAAGGATCAGAAGCATCTTCTCAAGAATATAC
ATAGAAGGAAACCTATACACAGCCACAGTCATCCACCTGCTTCGTGCGACTGATCAAGAAA
GAGCAGTGTGCAAGAGCAAATGGACAAGCTTTCACGTGAGAAAGCTGCAATTGAAGCTA
AGCTTTTAAAGTTCAAACAACAGAAGGTTGTAGCAAAGCATCAGTTTGAAGAAATGACTG
AGCATGTTGATGATATGGAGAATAGGCAGAGAAGAGCTGCTGAATTTTTTGGAAACTGCGA
TTCGGAATCCTACTTTTGTAAAGAAATTTTGGTAAGAAAGTCGAGCAGTTGGATATTTTCAG
CTTACAACAAAAAGCGAAGGCTCCCTGAAGTTGAGCAATCAAAGCCACCTTCAGAAGATT
CTCATCTGGATAATAGTAGTGGTAGCTCGAGACGCGAGTCTGGAAACATTTTTTCATCAA
ATTTCTCTAATAAATTCGCGACTAGAGCTTTCTCCAGCTGATTGATATGAACATGGTTT
CACACAGTATACAAAGTTCCAATGAAGAAGGTGCGAGTCCCAAAGGGATACTGTCCAGGAG
GTGATCCAAATACTACACTAACAAAAAGAGAAGGCTTACCATTGTCACCTGAAGCTCTAG
AGCTTGCGGATACCGGGACATGCCCGAGGAGATTACTGTAAATGATAATAACAAGGTGG
AGACCTTGCGAGCAGAGGCTAACTTCTTCAGAGGAGACTGATGGTAGCTTTTCATGTCATT
TAAATCTAACCTTGCTTCTGCTCCGTTACCGGACAAAACAGCTTCACAGATAGCTAAGA
CGACTCTTAAAGTCAGGAGTTAACTTTAACTCAATAGAAACAAGTGCAAGTGAGAAAA
ATCGGGGTAGACAAGAGATTGCAGTTGGAGGTAGCCAAAGCAAATGCAGCTCCTCCAGCAA
GAGTGAATGATGATTCTGGGAACAGTTCTTAACAGAAAGGCCAGGGTCTTCAGATAATG
AGGAGGCAAGTTTCGACTTATAGAGGTAACCCATACGAAGAGCAAGAGGAGAAAAGAAACG
GGAGTATGATGTTTACGTAATACAAAGAATATCGAGCAGCTGACCTTATAAACTATTTTGGA
CGGTTACATCAACGAGAGTACGAAGTGAAGTTTGGTAAGAAAGTATGGGTGAGTAAGTAA
TGAAACATTGGACTGAAAAAGCGTAAGTAGCTTTGTTGTAAACACTTGCGTCTCTGTCTA
CACAAGTAATTGACTGTAAATGTAAGTGTACAGGATTAAATTGAATAAGCA

>G264 Amino Acid Sequence (domain in AA coordinates: 24-114)

MNGALGNSSASVSGGEGAGGPAPFLVKTYEMVDDSTDQIVSWSANNNFIVWNHAEFSR
LLLPTYFKHNNFSSFIQNLNTYGFRRKIDPERWEFLNDDFIKDQKHLKNIHRRKPIHSHS
HPPASSTDQERAVLQEQMDKLSREKAAIEAKLLKFKQKQKVVAKHQFEEMTEHVDDMENRQ
KKLLNFLETAIRNPTFVKNFGKKVEQLDISAYNKKRRLPEVEQSKPPSEDSHLDNSSGSS
RRESGNIFHQNFNKLRLLELSPADSDMNMVSHSIQSSNEEGASPKGILSGGDPNTTLTKR
EGLPFAPEALELADTGTCPRRLLLNDNTRVETLQQLTSSEETDGSFCHLNLTLASAPL
PDKTASQIAKTTLSQELNFSIETSASEKNRGRQEIAVGGSQANAAPPARVNDVFWEQF
LTERPGSSDNEEASSTYRGNPYEEQEERNGSMMLRNNTKNIEQLTL*

>G32 (101..736)

AACACACATTCCTCTCTCTTCTTCAACTAGAAAAAGATAGATATATCGGACATTTATTG
ATCTGTGTATGCATAAAGGTATAGTATCATTATTAGAAAGATGAACACAACATCATCAA
GAGCAAGAAGAAGCAAGACGATCAGGTTGGTACAAGGTTTCTTGGGGTGAGAAGAAGGCC
TTGGGGAAGATACGCAGCTGAGATTAGAGACCAACTACGAAGGAGCGTCACTGGCTTGG
CACTTTTCGATACGGCGGAAGAAGCTGCCTTGGCCTACGATAGAGCTGCTCGGTCCATGCG
TGGCACACGTGCCAGAACCAACTTTGTTTACTCAGACATGCCTCCTTCTCATCCGTCAC
CTCCATTGTTTCTCTCAGCATCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTCCTAGCAA
TGATCCTGTGTCATGATGATGATGTTTAAACCAATACTCATCCACTGACTCGCCAATGCT
TCAGCCTCATTGTGATCAAGTGACAGTTACATGTTTGGTGGCTCTCAATCTTCGAATTC

TTATTGCTATTCTAATGACAGTAGTAATGAGCTGCCTCCTCTCCCGAGCGACTTGTGCGAA
TTCGTGTTATAGCCAACCACAGTGGACCTGGACCGGTGACGACTACTCGTCTGAGTACGT
ACATAGTCCAAATGTTTCAGCAGAATGCCTCCGGTTTCTGACTCTTCCCTCAAGGTTTCAA
CTACTTTGGCTCCTAATTCTTCTCATCGTCCATATTTAATACCTTCTCATTTGTACCT
TTTCCTTCTTCTTCTTTTTTGGGTTTATCTATGTTTCGCCGTCTTGATCTCTGCCTATG
TGATCAAAGTGACTGTTTGTCAATAGTTTTTCAATAACAAGTTATCATTTGTATCTTGAA
AAAAAAAAAAAA

>G32 Amino Acid Sequence (domain in aa coordinates: 17-84)
MNTTSSKSKKKQDDQVGTRFLGVRRRPWGRYAAEIRDPTTKERHWLGTFTAEAAALAYD
RAARSMRGTRARTNFVYSMDMPSSSVTSIVSPDDPPPPPPPPAPPSPNDPVDYMMMFNQYS
STDSPMLQPHCDQVDSYMFSGSQSSNSYCYSDSSNELPLPSDLSNSCYSPQOWTWTGD
DYSSEYVHSPMFSRMPFVSDSFPQGFNYFGS*

>G436 (1..2157)

ATGGATTTTACTCGCGATGACAACTCAAGTGATGAACGGGAAAATGATGTAGACGCCAAC
ACCAACAACCGTCACGAGAAGAAGGGTTACCATCGCCACACTAATGAACAAATTCATAGG
CTTGAAACGTATTTCAAGGAATGTCTCATCCAGACGAATTTAGCGACGTCTGTTGGGT
GAAGAAGTGAATCTGAAACCAAAAACAAATCAAATTTTGGTTTCAAACAAAAGAACTCAA
GCTAAGAGTCACAATGAAAAAGCAGACAATGCAGCGCTTAGGGCAGAAAATATTAAAGATT
AGACGTGAGAACGAATCAATGGAAGATGCACTGAATAATGTGGTTTGCCCTCCATGTGGT
GGTTCGTGGTCTCGGAGAGAGACCAACTTCGACATCTCAAAAACCTCCGTGCACAAAAC
GCTTATCTCAAAGATGAGTATGAAAGAGTCTCAAACACTCTAAAACAGTACGGAGGTCAC
TCAATGCATAACGTGAGGCCACACCTATCTCCATGGTCCATCAAACCATGCATCAACG
TCCAAGAACCGTCCAGCATTGTACGGAACCTCTTCTAACCGTCTCCCGAGCCTTCAAGC
ATATTTAGAGGACCATACCTCGTGGAACATGAACACCACCGCACCGCTCAGCCGCGA
AAGCCGCTGGAAATGCAGAAATTTCCAACCACTATCTCAACTGGAGAAAATGCAATGTTG
GAAGCAGCGGAAAAAGCGGTGTGAGAGGTTTGGAGCCTCATTCAAATGGATGATACAATG
TGGAAAAAGTCGTCTATTGATGATAGGCTCGTCATTGATCCAGGGCTCTATGAGAAATAT
TTTACTAAGACTAACACAAATGGTTCGTCTGAGTCTTCTAAAGATGTCTGGTGGTTCAA
ATGGATGCTGGAACTTGATCGACATCTTCTTAACTGCGGAGAAATGGGCGAGGCTTTTT
CCAACAATTGTGAACGAAGCTAAAACGATTACGCTCTTGGATTCCGTTGACCATCGAGGA
AAAACCTTTCTCAAGAGTGATTTATGAGCAACTGCACATACTGTACCATTTGGTGCCACCG
AGGGAATTTATGATCCTAAGGACTTGCCAACAAATGAAGACAATGTCTGGATGATTGCT
GATGTGTCGTGTCATCTCCCAAACATTGAGTTTGATCTTTCGTTTCCCATTGACCCAAA
CGTCCCTCAGGTGTGCTCATTCAGCCTTGCCCCACGGCTTCTCTAAGGTGACGTGGATA
GAGCATGTGGTAGTGAATGATAATAGAGTGCGGCCACATAAGCTTTACAGAGACCTCTTA
TACGGCGGCTTTGGCTACGGAGCTCGACGTTGGACCGTTACTCTTGAGAGGACGTGTGAG
AGGCTGATTTTCTCCACCTCCGTCCCTTGCCCAACAATGACAATCCCGGAGTTGTG
CAAACAATACGAGGCAGAAATAGCGTAATGCATTTGGGAGAAAAGAAATGTGAGGAACCTT
GCATGGATGATGAAAATGGTTAAACAACTCGACTTCTCGCCACAGTCTGAAACTAACAAAC
AGCGGAATTAGGATTGGGGTGGCGATAAACAAATGAGGCGGGTCAACCGCCCGGTCTCATT
GTCTGTGCTGGTTTCATCTTTATCCCTCCCTCTCCCTCTGTCCAAGTGTACGATTTCTCT
AAGAATCTGGAGGTTCTGCACCAAGTGGGACGTTCTGTGCCATGGGAATCCAGCGACTGAG
GCTGCTCGTTTCGTACCCGATCAAACCAAGGAACACTGTGTCTTTTCTCGAGCCTTCA
ATTAGGGATATTAATACTAAGCTAATGATACTCCAAGATAGCTTCAAAGATGCATTGGGA
GGAATGGTGGCCTACGCTCCAATGGATCTAAACACCGCCTGCGCTGCCATTTAGGCGAT
ATCGATCCTACCACATTCCAATCCTCCCTTCCGGTTTTATGATCTCCCGTGACGGCCGT
CCTTCCGAGGGCGAAGCCGAGGGTGGCAGCTATACACTCCTCACCCTGGCTTTCCAGATC
CTTGTCTCCCGTCCSAGTTACTCTCTGATACCAACCTGGAAGTTTCTGCCACCACAGTC
AATACCTTGATTAGCTCCACCGTTCAAAGGATCAAAGCCATGCTCAAGTGCGAATGA

>G436 Amino Acid Sequence (domain in AA coordinates: 22-85)
MDFTRRDSSDERENDVDANTNNRHEKKGYHRHTNEQIHRLETYFKBCPHPDEFQRRLLG
EELNLKPKQIKFWFQNKRTQAKSHNEKADNAALRAENIKIRRENESMEDALNNVCPPCG
GRGPGREDQLRHLQKLRAQNAAYLKDEYERVSNYLKQYGGHSMHNVEATPYLHGSPNHAST
SKNRPALYGTSSNRLPEPSSIFRGPYTRGNMNTTAPPQPRKPLEMQNFQPLSQLEKIAML
EAAEKAVSEVLSLIQMDDTMWKKSIDRLVIDPGLYEKYFTKTNTNGRPESKDVVVVQ
MDAGNLIDIFLTAEKWARLFPTIVNEAKTIHVLDSVDHRGKTFSRVIYEQLHILSPLVPP
REFMILRTCCQIEDNVWMIADV SCHLPNIEFDLSFPICTKRPSGVLIQALPHGFSKVTWI

EHVVVNDNRVRPHKLYRDLLYGGFGYGARRWTVTLERTCERLIFSTSVPALPNNDNPGVV
QTIRGRNSVMHLGERMLRNFAMMMKMNKLDSPQSETNNSGIRIGVRINNEAGQPPGLI
VCAGSSLSLPLPPVQVYDFLKNLEVRHQVDVLCHGNPATEAARFVTGSNPRNTVSFLEPS
IRDINTKLMILQDSFKDALGGMVAYAPMDLNTACAAISGDIDPTTIPILPSGFMISRDGR
PSEGEAEGGSYTLTLTVAFAQILVSGPSYSPDNTLEVSATTVNTLISSTVQRIKAMLKCE*

>G556 (50..1144)

CTTTTTTGAAGCCCTTTTGACACAAAAGACCAGAACAAGTTGAAGAAATATGAATACAAC
CTCGACACATTTTGTTCACCGAGAAGGTTTGAAGTTTACGAGCCTCTCAACCAAATCGG
TATGTGGGAAGAAAGTTTCAAGAACAATGGAGACATGTATACGCCTGGCTCTATCATAAT
CCCGACTAACGAAAAACCAGACAGCTTGTCTAGAGGATACTTCTCATGGGACAGAAGGAAC
TCCTCACAAAGTTTGACCAAGAGGCTTCCACATCTAGACATCCTGATAAGATACAGAGAAG
GCTAGCACAGAATCGAGAGGCAGCTAGGAAAAGTCGTTTGCAGCAAGAAAGCTTATGTTCA
GCAGCTAGAGACTAGCCGGTTAAAGCTAATTCATTTAGAGCAAGAACTCGATCGTGCTAG
ACAACAGGTTTCTTATGTGGGGAACGGAGTAGATACCAATGCTCTTAGTTTCTCAGATAA
CATGAGCTCAGGGATTGTTGCATTTGAGATGGAATATGGACATTGGGTGGAAGAACAAGAA
CAGGCAAAATATGTGAACCTAAGAACGGTTTACATGGACAAGTTAGTGATATAGAGCTTCG
TTCTCTAGTCGAGAATGCCATGAAACATTACTTTCAACTCTTCCGAATGAAGTCAGCCGC
TGCAAAAATCGATGTTTCTATGTCTATGTCGGAATGTGGAAAACCTCAGCAGAGCGGTT
TTTCTTGTGGATAGGCGGATTTAGACCCTCAGAGCTTCTCAAGGTTCTGTTACCGCATTT
TGATCCTTTGACGGATCAACAACCTTTTGGATGTATGTAATCTGAGGCAATCATGTCAACA
ATCAGAAGATGCGTTATCCCAAGGTATGGAGAACTGCAACATACATTAGCAGAGAGTGT
AGCAGCCGGGAAACTTGGTGAAGGAAGTTATATTCTCAAATGACTTGTGCTATGGAGAG
ATTGGAGGCTTTGGTCAGCTTTGTAAATCAAGCTGATCATCTGAGACATGAGACATTGCA
ACAGATGCATCGGATCTTAACCACGCGACAAGCGGCTAGAGGTTTGTAGCATTAGGGGA
GTATTTCCAAAGGCTTCGAGCTTTGAGTTTCGAGTTGGGCGGCTAGGCAACGTGAACCAAC
GTAATTAAGGTGTTTATAGTGTCAAGAAAGGTTTGAACCTTAACAATCAAGAATGGAGTT
TGCTGGTGAGTGGATTTTGGGTCAAGAACAAGAGCAATAACACAAGCTGCTGTGTGATG
ATGAATCTTGTCTTGGCGCTAAAGGAAATGTTTGAAGAAAGTTGTACATATGATCAGCAA
CGTAAAGTTTATAGCTTTTGTAGAAACCAACTTTTCGATGGTTGTTCTTTTTTTTTGTAT
GTAATATTATAGATAAGCTTGTGGTATATATGATTTTAATGTGACATTACGAACCTTGATT
TATAACCATGGTAAAAAT

>G556 Amino Acid Sequence (domain in AA coordinates: 83-143)

MNTTSTHFVPPRRFEVYEPLNQIGMWEESEFKNNGDMYTPGSIIPTNEKPDLSLSEDTSHG
TEGTPHKFDQEASTSRHPDKIQRRLAQNREAAKSRLRKKAYVQQLSRLKLIHLEQEL
DRAQQGFYVGNVDVTNALSFSNMSGIVAFEMEYGHWVEEQNRQICELRTVLHGQVSD
IELRSLVENAMKHYFQLFRMKSAAAKIDVFYVMSGMWKTSERFFLWIGGFRPSELLKVL
LPHPDPLTDQQLLDVNCNLRQSCQSEDALSQGMKQLQHTLAESVAAGKLGEYSYIPQMT
AMERLEALVSFVNQADHLRHETLQOMHRILTTRQAARGLLALGEYFQRLRALSSSWAARQ
REPT*

>G1420 (39..1238)

AAAGTATCATCTCATAGATTCCATCTTTTCTCTATTACATGGAGAAGAAAAAGAAGAGG
ATCATCATCATCAACAACAACAACAACAAGGAGATCAAGAACACAGAGACAAAGA
TCGAGCAAGAACAAGAACAAGAACAACAAGAAATCTCTCAAGCATCATCATCAAA
ACATGGCGAATCTAGTTACGTCTATCAGATCATCATCCGTTGGAGCTAGCTGGAAATCTCT
CAAGCATCTTCGATACCTCATCTTTACCTTTTCTTATTCTTATTTTGAAGATCACTCTT
CTAATAATCCTAATCTTTTCTTAGACTTGCTCCGACAAGATCATCAGTTTGCTTCTTCT
CTAATTCCTCTTCTTTTTCATTGATGCCCTTTCTCTCCCCAATAACAACAACAACCT
CTTTTTTTACGGAATTTGCCCTTACCTCAAGCTGAGTCATCAGAAGTCGTGAACACAACAC
CGAGTCTCTCAAACTCAACCTCAGTCTCATCTTCTCCTCAACGAAGCTGCAAAATGATAACA
ACAGTGGTAAAGAAGTTACTGTAAAGATCAAGAAGAAGGAGATCAACAACAAGAGCAAA
AGGGTACTAAGCCACAGTTGAAGGCAAGAAGAAGAATCAAAAGAAAGCTAGAGAAGCTA
GGTTTGGCTTTCTGACGAAGAGCGATATTGATAATCTTGACGACGGTTATAGGTGGAGAA
AATACGGCCAAAAGCTGTCAAAAACAGTCTTATCCAGAAGCTATTACCGTTGCACCA
CAGTGGGTTGCGGAGTGAAGAAGAGAGTGAGAGATCTCCGATGATCCTTCGATCGTCA
TGACAACCTACGAAGTCAAGATACCATCTCTTCCCATGACGCCACGTGGACACATCG
GAATGCTCAGCTCAGCAATCCTAGACCAGGTGCAACCACGCGTCATCATCATCATTCT
CCATCCCTCAGCCACGTTACTTGCTGACTCAACATCACCAGCCCTACAACATGTACAACA

ACAACCTCTCTAAGTATGATCAATAGAAGATCATCCGATGGCACTTTTCGTAAATCCAGGTC
CATCATCATCATTCCTCCGGCTTTGGTTATGATATGTCTCAAGCTTCTACTTCAACTTCTT
CTTCCATTAGAGATCATGGATTGCTTCAAGATATTCTTCTTCGAGATCAGATCCGATA
CTATTAACACTCAAACCAATGAAGAGAATAAGAAATGAAGAAGTTTTTTTCCCGGGGCA
ATTGTTTTTTTCTTTAGGCCGGATCCGGTAGGTAGGTTTCATGAGC
>G1420 Amino Acid Sequence (domain in AA coordinates: 221-280)
MEKKKEEDHHHQQQQQQQKEIKNTETKIEQEQQEQKQEQISQASSSSNMANLVTSSDHHF
LELAGNLSSIFDTSSLPPFYSYFEDHSSNNPNSFLDLLRQDHQFASSNSSFSDAFPL
PNNNNNTSFFDLPPLQAESSEVVNTTPTSPNSTSVSSSSNEAANDNNSGKEVTVKDQEE
GDQQEQKGTQKPKLAKKKKNQKKAREARFAFLTKSDIDNLDGGRWRKYQKAVKNSPYP
RSYYRCTTVGCGVKRVERSSDDPSIVMTTYEGQHTHPFPMTPRGHIGMLTSPILDHGAT
TASSSSFSIPQPRYLLTQHHPYNMNNNSLSMINRRSSDGTFTVNPSPSSSPGFGYDMS
QASTSTSSSIRDHGLLDILPSQIRSDTINTQTNEENKK*
>G1412 (115..1008)
CCCACGCGTCCGCCCACGCGTCCGAAACAAAAACATATAATTTGGGTTTITAGAGTTCGA
AACTTGAAATCTTTTTTTTTTTGGTTGCTGAGGAATCGAAGTAGAAGAGTATAAATGGGT
GTTAGAGAGAAAGATCCGTTAGCCAGTTGAGTTTGCCACCAGGTTTATGATTTTATCCG
ACAGATGAAGAGCTTCTTGTTCAGTATCTATGTGCGAAAGTTGCAGGCTATCATTTCTCT
CTCCAGGTCATCGGAGACATCGATCTCTACAAGTTTCGATCCTTGGGATTGCGCAAGTAAG
GCTTTGTTTGGAGAGAAGGAATGGTATTTCTTTAGCCCAAGAGATCGGAAATATCCGAAC
GGGTCAAGACCAATAGAGTAGCCGGTTCGGTTATTGAAAGCAACGGGTACTGACAAA
ATTATCACGGCGGATGGTCTCGTGTTCGGGATTAAAAAAGCTCTGGTCTTTACGCCGGA
AAAGCTCCCAAAGGCACTAAACCAACTGGATTATGCACGAGTATCGCTTAATAGAACAT
TCTCGTAGCCATGGAAGCTCCAAGTTGGATGATTGGGTGTTGTGTCGAATTTACAAGAAA
ACATCTGGATCTCAGAGACAAGCTGTTACTCTGTTCAAGCTTGTCGTGAAGAGCATAGC
ACGAATGGGTGCTCATCGTCTTCTTCATCACAGCTTGACGACGTTCTTGATTCTGTTCCCG
GAGATAAAAGACAGTCTTTTAATCTTCTCGGATGAATTCGCTCAGGACGATTCTTAAC
GGGAACTTTGATTGGGCTAGCTTGGCAGGTCTTAATCCAATTCAGAGCTAGCTCCGACC
AATGGATTACCGAGTTACGGTGGTTACGATGCGTTTCGAGCGGCGGAAGGTGAGGCGGAG
AGTGGGCATGTGAATCGGCAGCAGAACTCGAGCGGGTTGACTCAGAGTTTCGGGTACAGC
TCGAGTGGGTTTGGTGTTCGGGTCAAACATTGAGTTTAGGCAATGAGAGAGATGTGAA
GTTACTGATGGGTGAGAAAAAGTAAAAAAGAACTGGAGATAGTAGAGTGGCAATTGATG
TAAATAATAGGGATTATATAGGGGCTTTTACCGATTGCGTGAGGCTTAGGATTCCCCAAA
GGAAAAAGGCTCGACTGGGACTAGTTTGTATCCAACCTGACGGCCCCCAATGTGTAATG
TTTCTCAACGAGAGAAAAATAATGGTTACCAATATTTTCCAAAAAAGAAAAA
>G1412 Amino Acid Sequence (domain in AA coordinates: 17-159)
MGVREKDLPLAQLSLPPGFRFYPTDEELLVQYLCKRVAGYHFSLQVIGDIDLYKFPDWDLP
SKALFGEKEWYFFSPRDKYPNGSRPNRVAGSGYWKATGTDKIITADGRRVGIIKALVYF
AGKAPKGTKTNWIMHEYRLIEHSRSHGSSKLDWVLCRIYKKTSGSQRAVTPVQACREE
HSTNGSSSSSSQLDDVLDSPFEIKDQSFNLPRMNSLRILNGNFDWASLAGLNPIPELA
PTNGLPSYGGYDAFRAEAGEAESGHVNRQNSSGLTQSFYSSSGFVSGQTFEFRQ*
>G738 (1..885)
ATGGACCATCATCAGTATCATCATCATGATCAATACCAACATCAGATGATGACTAGTACT
AACAATAATTCCTATAACACCATCGTCACAACACAACCACCACCAACAACAACAATG
GATTCACAACACAGCAACAACATATGATAATGGATGACGAGAAGAAGTTGATGACGACAATG
AGCACTAGGCCCGCAAGAACCAAGAACTGTCCAAGATGCAACTCAAGCAACACCAAGTTT
TGTTATTACAACAACACAGCTTAGCACAGCCTAGGTACTTGTGTAAGTCTTGTTCGGAGA
TATTGGACTGAAGGTGGCTCTCTCCGTAACGTCCCCGTAGGCGGAGGTTCTAGAAAGAAC
AAGAAGCTTCCATTTCTTAATTCCTTCTTCTTCTTCCACCAAGAACCTCCCGGATCTC
AACCCTCCTTTCGTCTTCACATCATCAGCTTCATCATCAAACCTAGCAAGACGCATCAA
AACAATAATGACCTCAGCCTATCCTTCTCCTCCCCATGCAAGACAAGCGAGCTCAAGGG
CATTACGCTCATTTCACTGAGCAAGTTGTGACAGGAGGGCAGAACTGTCTTTTCCAAGCT
CCTATGGGAATGATTCACTTTCGTCAAGAGTATGATCATGAGCACCACCAAGAAATCTT
GGGTTTTTCATTAGACAGGAACGAGGAAGAGATTGGTAATCATGATAACTTCGTTGTTAAT
GAGGAAGGAAGTAAGATGATGTATCCTTATGGAGATCATGAAGACCGTCAACAACATCAC
CATGTGAGACACGATGATGGTAATAAGAAGAGAGAAGGTGGTTCAAGCAATGAGCTATGG
AGCGGAATCATCTAGGTGGTGATAGTGGTGACCAACATGGTGA

>G738 Amino Acid Sequence (domain in aa coordinates: 351-393)
MDHHQYHHHDQYQHQMSTNNNSYNTIVTTQPPPTTTTMDSTTATTMIMDDEKKLMTM
STRPQEPNRCPRCNSSNTKFCYNNYSLAQPRYLCKSCRRYWTEGGSLRNVPVGGGSRKN
KKLPFPNSSTSSSTKNLPDLNPPFVFTSSASSSNPSKTHQNNNDLSLSFSSPMQDKRAQG
HYGHFSEQVVTGGQNCFLQAPMGMIQFRQYDHEHPKKNLGFSLDRNEEEIGNHDFVFN
EEGSKMMPYDGHEDRQHHHVRHDDGNKKREGGSSNELWSGIILGDSGGPTW*

>G2426 (1..1038)

ATGGGCAGATCGCCATGTTGTGATAAGGCCGGGTGAAGAAAGGCCCTTGGACTCCAGAA
GAGGATCAGAAACTTTTGGCTTATATTGAAGAACATGGCCATGGAAGCTGGCGTCTTTG
CCTGAGAAAGCCGGTCTCCAAAGGTGTGGAAGAGTTGCAGACTCAGATGGACTAATAC
CTAAGACCTGACATCAAGAGAGGCAAATCACTGTACAAGAAAGACAAACCATCATTCAA
CTCCACGCTCTCCTCGGAAACAGGTGGTCAGCGATTGCAACTCATTTACCAAAGAGGACA
GACAACGAGATCAAGAACTACTGGAACACACACTTGAAGAAACGTCTGATCAAAATGGGG
ATAGATCCAGTGAATCACAAGCACAAAACGAGACTCTTTCGTCTTCCACAGGACAATCA
AAGAACGACGACGCTTAGTCATATGGCTCAATGGGAGAGTGCAAGACTCGACGCTGAA
GCAAGGCTAGCTAGAGAATCAAAGCTTCTCCATTTACAGCATTACCAAAACAATAACAAC
CTTAACAAATCAGCAGCTCCTCAACAACATTGCTTCACTCAAAAACATCAACAACTGG
ACTAAACCAAACCAAGGAAACGAGACCAACAGCTTGAATCTCCGACATCGACGGTGACA
TTCTCTGAGAATCTTCTGATGCCTTTAGGAATCCCTACGGATAGCAGCAGAAATAGAAAC
AATAACAACATGAGTCTCGGCGATGATTGAATTGGCCGTATCTTCGTCAACCTCCTCC
GATGTGAGTCTGGTCAAAGAACATGAACACGACTGGATTAGGCAGATCAACTGTGGTAGT
GGAGGAATAGGAGAAGGATTACAGAGTCTATTGATCGGTGATTGCGGTCGGCCGGGGTTA
CCCACCGGGAACGAAGCGACGGCGGGCGTGGGAATGAGAGTGAGTATACTACTAT
GAGGATAACAAGAATTACTGGAATAGCATTCTCAACTTGGTTGATTCTTCACCGTCCGAT
TCCGCGACGATGTTCTGA

>G2426 Amino Acid Sequence (conserved domain in AA coordinates:14-114)

MGRSPCCDKAGLKKGPWTPEEDQKLLAYIEEHGHGSWRSLEKAGLQRCGKSCRLRWNTY
LRPDIKRKFVQEEQTIIQLHALLGNRWSAIATHLPKRTDNEIKNYWNTHLKKRLIKMG
IDPVTHKHKNETLSSSTGQSKNAATLSHMAQWESARLDAEARLARESCLLHLQHYQNNNN
LNKSAAPQQHCFQKTSTNWTNPKPNQNGDQQLSPTSTVTFSENLLMPLGIPTDSSNRN
NNNNESSAMIELAVSSSTSDVSLVKEHEHDWIRQINCGSGGIGEGFTSLIGDSVGRGL
PTGKNEATAGVNESEYNYEDNKNYWNLSILNLVDSSPSDSATMF*

>G1524 (1..825)

ATGGGGAGAACTAAGGAGCAGGCAACATTAACCTCGGTATCCACCCTGTCTTAGGAATCCT
GCTAAATTCAATGATATAAACAAGCACTCCAGGAAAAAGGATATGGTAAGGCTCTGAAA
AGAAAAACCTTGGACGGGTGTGACATGCCCTGTCTGTCTTGAGGTTCTCACAACCTCGGTC
GTCTCTCTTTGTTTCATCTTACCACAAAGGATGCCGTCCGTACATGTGTGCCACGGGAAAC
CGTTTCTCAAATTGTCTAGAGCAGTACAAAAAGGCATATGCCAAGGATGAGAAAAGTGAC
AAACCGCCAGAGCTATTGTGCCCGCTTTGTAGGGGTGAGGTGAAAGGCTGGACCGTTGTG
GAAAAGGAACGTAAGTATCTGAATTCTAAGAAAAGGTCATGCATGAACGACGAGTGTGTTG
TTTTATGGAAGCTATAGACAGCTCAAGAAGCATGTTAAGGAGAACCATCCGAGAGCCAAG
CCAAGAGCCATAGACCTGTGTCTGGAGGCGAAATGGAAGAAGCTTGAGGTTGAGAGGGAG
AGGAGTGATGTAATCAGCACAGTCATGTCTCAACACCTGGGGCTATGGTATTTGGAGAC
TATGTGATTGAGCCATACAATGGTTATGATCATCAAGATGACAGTGACGATTACAGTGAT
TCGTCCGATGACGAAATGGAAGGTGGGGTATTTCGAGCTTGGAGCATTCGACCTGGGCCGT
CTTCAACCGCGTTCCGCTGCCATCTCAAGCCGGGGAATTCCGCGTATGATCATAAGGAAC
CGGTGGGCTCGAAGCAGAGGTGCGAGCAGAAGGCGACAAACATAA

>G1524 Amino Acid Sequence (conserved domain in AA coordinates:49-110)

MGRTEQATLTRYPPCPNPAKFNDINKALQEKYKALKRKPWTGVTCPCVLEVPHNSV
VLLCSSYHKGCRPYMCATGNRFSNCLEQYKKAYAKDEKSDKPPPELLCPLCRGQVKGWTVV
EKERKYLNSKKRSCMNDECLFYGSYRQLKKHVKENHPRAPRAIDPVLEAKWKLEVERE
RSDVISTVMSSTPGAMVFGDYVIEPYNGYDHQDDSDSDSDEMEGGVFELGAFDLGR
LQPRSAAISSRGIRMIIRNRWARSRGASRRRQT*

>G1243 (1..3174)

ATGGCGAGAAATTCGAATTCGATGAGGCTTTCTCGTCAGAGGAGGAAGAAGAGCGGGTT
AAGGATAATGAAGAAGAAGATGAGGAGGAGCTCGAGGCTGTTGCTCTTCTTGCTCC
GACGATGACGAAGTAGCCCGCCGACGAATCACCAGTCTCCGACGAGAGGCTGCTCCC

GTAGAAGATGATTACGAGGACGAAGAAGATGAGGAAAAAGCTGAAATCAGCAAACGTGAG
AAAGCCAGACTTAAAGAGATGCAGAAGTTGAAGAAGCAGAAGATTCAAGAGATGCTGGAG
TCGCAGAATGCTTCCATTGACGCGGATATGAACAATAAGGGAAAAGGGAGACTGAAGTAT
CTTCTGCAGCAAACCTGAGTTATTTGCCCACTTTGCTAAAAGTGATGGATCTTCTTCTCAG
AAGAAGGCAAAAGGAAGGGGACGTCATGCTTCCAAAATAACTGAAGAGGAGGAAGACGAA
GAGTATCTAAAGGAAGAAGAGGATGGCTTAACTGGATCTGGAAACACACGGTTACTCACA
CAGCCCTCTTGTATTCAAGGGAAGATGAGAGATTACCAATTAGCTGGTTTGAAGTGGCTC
ATTCTGCTTTTATGAGAATGGCATAAATGGAATTCTTGCTGATGAAATGGGTCTGGGGAAG
ACGCTTCAAACGATTTCTTTGTTGGCATATCTTCATGAATACAGGGGAATCAATGGTCCC
CATATGGTGGTTGCTCCAAAATCAACACTTGGTAATTGGATGAACGAAATTCGCCGGTTT
TGCTCTGTCTACGTGCTGTGAAGTTCTTGTAATCTTGAGGAGAGGAGACATATTCGA
GAAGACCTGCTAGTTGCTGGGAAATTTGATATTTGTGTCAAGCTTTGAGATGGCCATC
AAAGAGAAGACAGCACTTCGTCGGTTTAGCTGGCGTTATATTATCATTGATGAAGCGCAT
CGAATCAAGAACGAGAATTCAC'TCCTTTCTAAAACCATGAGACTTTTTAGCACCAATTAT
CGGCTTCTTATCACGGGGACCCCCCTTCAGAATAATCTCCATGAAGTGTGGCTCTTCTA
AATTTTCTTCTGCCTGAGATTTTGTAGTTCAGCAGAGACTTTTGATGAATGGTTTCAAATT
TCTGGTGAGAAATGACCAGCAAGAAGTTGTGCAACAACCTGCACAAGGTTCTTCGACCATT
CTTCTTCGAAGACTAAAGTCAGATGTTGAGAAAGGTTTGCACCGAAGAAGGAGACCATA
CTTAAAGTTGGTATGTCTCAGATGCAAAAGCAATACTACAAGGCTTTACTGCAGAAGGAT
CTTGAAGCGTTAATGCTGGTGGAGAACGCAACGCTCTGCTAAACATTGCAATGCAACTG
CGTAAATGCTGCAATCACCCCTATCTCTCCAGGGTGCAGAACCTGGTCCCCCATATACC
ACAGGAGATCACCTTATAACAAATGCTGGTAAGATGGTTCTCTTGGAATAAATTGCTTCCT
AAGTTGAAAGAACGTGATTCAAGGGTGTGATATTTTCTCAGATGACAAGACTTTTGGAT
ATTCTTGAGGACTATTTAATGTATCGTGGTTACTTGTATTGCCGTATTGATGGAACACT
GGTGGTGACGAACGAGATGCCTCCATAGAAGCCTACAACAAGCCAGGAAGTGAGAAATTT
GTTTTCTTGTATCTACTAGAGCTGGAGGGCTTGGTATCAATCTTGCTACTGCAGATGTT
GTGATCCTTTACGATAGTGATTGGAACCCACAAGTCGACTTGCAAGCTCAGGATCGTGCC
CATAGGATTGGTCAAAAAAAGAAGTTCAAGTGTTCGATTCTGCACTGAGTCTGCTATT
GAGGAGAAAGTGATTGAAAGAGCTTACAGAAGTTAGCACTTGATGCTCTGGTTATTCAA
CAAGGGAGATTGGCAGAACAGAAAAGTAAGTCTGTCAATAAGGATGAGTTGCTTCAAATG
GTAAGATATGGTGCTGAGATGGTGTTCAGTTCTAAAGATAGCACAAATCACAGACGAGGAT
ATTGATAGAATCATTGCCAAAGGAGAAGAGGCAACAGCTGAACCTTGATGCTAAGATGAAG
AAATTCACAGAAGATGCTATACAGTTTAAATGGATGACAGTGTGACTTCTATGATTTT
GATGATGACAATAAGGATGAAAACAAGCTCGATTTTAAAGATTGTAAGCGACAATTGG
AATGATCCCCCAAGCGGGAGAGAAAAGCGCAACTACTCTGAATCTGAGTACTTTAAGCAA
ACATTGCGGCAAGGTGCTCCAGCTAAACCTAAAGAGCCTAGAATTCGCGCATGCCCCAG
TTGCACGATTTCCAGTTCTTTAACATTAGAGATTGACCGAGTTGTATGAAAAGGAAGTA
CGTTATCTCATGCAAAACACATCAGAAAATCAGTTGAAAGACACAATTGATGTTGAAGAA
CCAGAAGGTGGGGATCCCTTAACTACTGAAGAAGTAGAAGAAAAGGAGGATTATTGGAG
GAGGGTTTCTCAACATGGAGCAGAAGAGATTTTAACTTTCTCAGGGCTTGTGAGAAG
TATGGCCGCAACGACATAAAAAGCATTGCCTCTGAGATGGAAGGGAAAACAGAGGAAGAA
GTTGAAAGATATGCCAAAGTATTTAAAGAGCGGTACAAGGAGCTGAACGACTATGATAGA
ATCATTAAAGAACATTGAGAGGGGAGAGGCAAGGATCTCTAGGAAAGACGAAATCATGAAG
GCCATAGGGAAGAAACTGGATCGCTACAGAAACCCTTGGCTGGAAGTGAAGATTCAATAT
GGTCAGAACAAAGGCAAGCTGTACAATGAAGAGTGTGACCGTTTCATGATCTGCATGATT
CACAACTTGGTTATGGGAATTGGGATGAGCTAAAGGCAGCATTTAGGACATCGTCTGTG
TTCAGGTTTGACTGGTTTGTGAAATCCCGCAGAGTCAGGAACCTTGCAAGAGATGCGAC
ACTCTGATTGACTGATCGAGAAAGAGAACCAGGAGTTTGATGAAAGAGAGAGGCAAGCC
CGCAAAGAGAAGAAGCTCGCGAAGAGTGCAACACCATCAAAGCGACCTTTAGGAAGACAA
GCAAGTGAGAGTCTTTCATCGACGAAGAAGCGGAAGCACCTGTCTGATGAGATGA
>G1243 Amino Acid Sequence (domain in AA coordinates: 216-609)
MARNSNSDEAFSSEEEERVKNDEEEDEEELEAVARSSGSDDEVAADSPVSDGEAAP
VEDDYEDDEDEEKAEISKREKARLKEMQKLKQKIQEMLESQNASIDADMNNKGKGRLLKY
LLQOTELFAHFAKSDGSSSQKKAKGRGRHASKITEEEDEEYLKEEDGLTSGNTRLLT
QPSCIQGMKRDYQLAGLNLIRLYENGINGILADEMGLGKTLQITISLLAYLHEYRGINGP
HMOVAPKSTLGNWMNEIRRFPCVLRVAVKFLGNPEERRHIREDLLVAGKFDICVTSFEMAI
KEKTALRRFSWRYIIIDEAHRICKNENSLLSKTMRLFSTNYRLLITGTPLQNNLHELWALL

>G631 (190..1461)

>G631 Amino Acid Sequence (domain in AA coordinates: TBD)

MDKEKSPAPPPSGGLPPPSGRYSAFSPNGSSFAMKAESSFPPLTPSGSNSDDANRFSHDI
SRMPDNPPKNLGHRRAHSEILTLPPDDLSDLSDLGVVGADGPSFSDDTDEDLLMYLDM
KFNSATSTSQMGEPSEPTWRNELASTSNLQSTPGSSSERPRIHQHSQSMGSTTIKPE
MLMSGNEDVSGVDSKKAISAAKLSSELALIDPKRAKRIWANRQSAARSKERKMRYIAELER
KVQTLQTEATSLSAQLTLLQRDTNGLGVENNELKLRVQTMEQQVHLQDALNDALKEEVQH
LKVTLTGQGPSNGTSMNYGSFGSNQQFYPNNQSMHTILAAQQLQQLQIQSQKQQQQQQHQ
QQQQQQQQQFHQQQQQLYQLQQQRLQQEQQSGASELRRPMPSPGQKESVTSPPDRETPL
TKD*

>G1909 (1..828)

ATGGGTGGATCGATGGCGGAGAGAGCAAGGCAGGCCAACATTCTCTCCACTAGCGGGACCC
CTAAAGTGCTCTCGATGCGACTCCAGCAACACTAAGTTCTGTTACTACAACACTATAAC
CTCACTCAGCTCGTCACTTCTGCAAAGGTTGCCGTGCTACTGGACACAAGGGGGCGCC
CTGAGAAACGTCCCTGTAGTGGAGCCGTGCCGGAGGAATAACAAGAAGGGCAAAAATGGA
AATTAAAATCTTCTTCTTCTTCTGCTCCAAACAGTCTTCTCGGTCAACGCTCAAGTCTCT
AGCTCAGGACGACCTAAGACAAATCATCAGTTCCCTTTTTCACCAACTCTTTTCAATCTC

ACTCAACTCGGAGGTATTGGTTTGAACCTTAGCCGCTACTAATGGCAACAACCAAGCTCAC
CAGATCGGTTCCAGTTTGATGATGAGCGATCTAGGGTTTCTCCATGGACGAAATACTTCA
ACTCCGATGACGGGAAACATTCATGAAAACAACAATAATAACAATGAAAACAACCTA
ATGGCATCCGTTGGATCTTTGAGCCCTTTGCTCTCTTCGATCCAACGACGGGGCTATAC
GCTTTCAGAACGACGGTAATATCGGGAACAACGTTGGGATATCTGGTTCTTCTACTTCC
ATGGTTGATTCTAGGGTTTATCAGACGCCTCCGGTGAAGATGGAAGAACAACCTAATTTG
GCTAACTTGTCTAGACCGGTCTCCGGTTTGACGTCTCCTGGGAATCAAACAAATCAGTAC
TTTTGGCCTGGTTCCGATTTCTCGGGTCCTTCTAATGATCTCTTGTA

>G1909 Amino Acid Sequence (conserved domain in AA coordinates:23-51)

MGGMAERARQANIPPLAGPLKPCRDSSNTKFCYNNYNLTQPRHFCKGCRRYWTQGA
LRNVPVGGGCRNNKKGKNGNLKSSSSSSKQSSSVNAQSPSSGQLRTNHQFPFSPPLYNL
TQLGGIGLNLAAATNGNNQAHQIGSSLMSDLGLFHLGRNTSTPMTGNIHENNNNNNNNNL
MASVGSLSPPFALFDPPTGLYAFQNDGNIGNNVGISGSSSTSMVDSRVYQTPPVKMEEQPNL
ANLSRPVSGLTSPGNQTNQYFWPGSDFSGPSNDLL*

>G1663 (64..630)

TTCTCTCTGTGAATCCTTGTTTCATCGTCACTGAAATTAGTTTACAAAATCGACGAATTTCG
GAGATGATTTTTTCAGAAATGTGTGCAGAAATGAGTCCAACCTTCAACGCTATAGCTTCCGAA
TCGCGTTTCCCAAACGCAGTTCGGTGTTCGAAATCCTCCTCGAGCGGCGGCGGATGTATC
TCCGCCAGGACTAAAGACCGTCACACGAAGGTTAACGGACGAAGCCGTCGAGTTACGATG
CCGGCTCTCGCCGCCGCTAGGATTTTCCAGTTAACCGGTGAGCTCGGTACAAAACCTGAA
GGAGAAACCATCGAATGGCTTCTTAGTCAAGCTGAACCGTCGATTATTGCCGCCACTGGC
TACGGGACTAAGCTCATTTTCAATTTGGGTTGATGTTGCGGCGGACGATTCTCGTCTGTCG
TCGTGATGACGTGCGCGCAACGCAACGCAACGCCACAATCGCCGAGTTGTAGGTTG
GATCTTTGTTCAGCCAATCGGAATTCAGTATCCGGTGAATGGTTACAGTCATATGCCGTTT
ACAGCGATGCTTTTAGAGCCGATGACCACGACGCGGAATCTGAGGTTGAGATCGCGGAG
GAGGAGGAACGTAGACGCCGTCACCATTAGTAAAATTAGGCTTTTGATTTAGAGTGTTAA
AATTAGGATTTTAAAGTTTAGGAGGTAACAGATAAGGATAATT

>G1663 Amino Acid Sequence (domain in AA coordinates: TBD)

MIFQNVCRNESNFNAIASESRSQTFGVSKSSSSGGGCGISARTKDRHTKVNRSRRVTMP
ALAAARIFQLTRELGHKTEGETIEWLLSQAEPSSIIAATGYGKTLISNWVDVAADDSSSSS
SMTSPQTQTPTQSPSCRLDLCPIGIQYPVNGYSHMPFTAMLLPMTTTAESEVEIAEE
EERRRRHH*

>G1231 (103..870)

CAAACCCAAATTTCTCAGCGCCGGTCAAATACTTGTCTCTCTCTCTCTCTCTTTTAC
TCTTGTCTTGTCTCTCTTGAAGCTGTTTGTCTGTAAAGATGGAAGCAGGTGGCGCG
TACAATCCACGCACTGTTGAAGAGGTGTTTAGGGATTTTAAAGGGTCGTAGAGCTGGCATG
ATTAAGGCTTTAACCAGTGATGTTTCAAGAGTTTTCGACTTTGTGATCCCCGAAAAGGAG
AACCTTTGCTTTTACGACATCCAAATGAGCACTGGGAAGTGAATTTGCCAGCTGAAGAG
GTTCTCTCTGAGCTCCCAGAGCCTGTCTTGGGTATCAATTTTGCCAGAGACGGGATGGCG
GAAAAGGATTGGTTGTCCCTTGTGCTGTCCACAGTGATGCTTGGCTTCTTGTGTTGCT
TTCTTTTTTGGAGCCAGGTTTGGATTTGACAAAGCTGATAGGAAGAGGCTTTTCAATATG
GTGAATGACCTCCCAACAATCTTTGAGGTTGTAGCTGGCACTGCTAAGAAACAAGGAAAA
GATAAGTCCCTCTGTTTCCAACAACAGCAGCAACAGATCCAAATCAAGCTCCAAGCGAGGA
TCTGAATCCCGTGCCCAAGTTCTCAAAGCCGAGCCCAAAGATGATGAGGAGGAGGAAGAG
GAAGGTGTGGAAGAGGAGGATGAGGATGAGCAAGGTGAACACAGTGTGGAGCATGTGGT
GAGAGCTATGCAGCTGATGAGTTCTGGATTGCTGTGACCTCTGTGAGATGTGGTTTCAT
GGAAAGTGTGTTAAGATAACACCAGCAAGAGCTGAGCACATCAAGCAATACAAGTGCCCT
TCTTGCAGCAACAAAAGGGCTCGTTTCTAAATTTGTTGACCGCTCGCTTCTGTGTATCTA
CCTTTGCATATGATGATGAACAGCTTAACTGTTTGGTTTAGATCAGATTTGTTCATATGGA
TTTGGTAATTTAGGAAGACATTTTAGTTTTCATTGTTACATTTTGGCGATTGAAGGGA
TAACTCTTTGTTTAGGGTAATGATCTTTGCTCTGTTTATGTTTATTATTAACATTC
TTCAAACCTCAATCAAAAGTATTTTGGTTAGTCTTAAAA

>G1231 Amino Acid Sequence (domain in AA coordinates: TBD)

MEAGGAYNPRTVEEVFRDFKGRRAGMIKALTVDVQEPFRLCDPEKENLCLYGHNPNEHVEV
NLPAAEEVPELPEPVLGINFARDGMAEKDWLSLVAVHSDAWLLAVAFFFGARFGFDKADR
KRLFNMVNDLPTIFEVVGATAKKQKDKSSVSNNSSNRKSSSKRGSESRAKFSKPEPKD
DEEEEBEGVEEDEDQGETQCGACGESYAADEFWICDLCCEMWFHGKCVKITPARAEHI

KQYKCPSCSNKRARS*

>G227 (21..983)

GTACCGTCGACGATCCGGCGATGTCAAACCCGACCCGTAAGAATATGGAGAGGATTAAAG
GTCCATGGAGTCCAGAAGAAGATGATCTGTTGCAGAGGCTTGTTCAGAAACATGGTCCGA
GGAAGTGGTCTTTGATTAGCAAATCAATCCCTGGACGTTCCGGCAAATCTTGTCTCTCC
GGTGGTGTAAACGAGTATCTCCGGAGGTAGAGCACCGTGCTTTTTCGAGGAAGAAGACG
AGACGATTATTCGAGCTCACGCTCGGTTTGGTAACAAGTGGGCTACGATCTCTCGTCTTC
TCAATGGACGAACCGATAACGCTATCAAGAATCATTGGAACTCGACGCTGAAGCGAAAAT
GCAGCGTCGAAGGGCAAAGTTGTGATTTTGGTGGTAATGGAGGGTATGATGGTAATTTAG
GAGAAGAGCAACCGTTGAAACGTACGGCGAGTGGTGGTGGTGGTGTCTCGACTGGCTTGT
ATATGAGTCCCGGAAGTCCATCGGGATCTGACGTCAGCGAGCAATCTAGTGGTGGTGAC
ACGTGTTTAAACCAACGGTTAGATCTGAGGTTACAGCGTCATCGTCTGGTGAAGATCCTC
CAACTTATCTTAGTTTGTCTCTTCTTGGACTGACGAGACGGTTCGAGTCAACGAGCCGG
TTCACTTAACAGAAATACGGTTATGGACGGTGGTTATACGGCGGAGCTGTTTCCGGTTA
GAAAGGAAGAGCAAGTGAAGTAGAAGAAGAAGCGAAGGGGATATCTGGTGGATTCTG
GTGGTGGTTCATGACGGTGGTTCAGGAGATGATAAGGACGGAGGTGAGGAGTTACATGG
CGGATTTACAGCGAGGAAACGTGGTGGTAGTAGTTCTGGCGGCGAGGTGGCGGTTCTGT
GTATGCCACAAAGTGTAAACAGCCGTCTGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
TCGGAATTGGGAAGATGGAGTAGGCGGCC

>G227 Amino Acid Sequence (domain in AA coordinates: 13-112)
MSNPTKRKNMERIKGPWSPEEDLLQLVQKHGPRNWSLISKSIPIGRSGKSCRLRWCNQLS
PEVEHRAFSQEEDETIIIRAHARFGNKWATISRLNNGRTDINAIKNHWNSTLKRKCSVEGQS
CDFGGNGGYDGNLGEELQLKRTASGGGGVSTGLYMSPGSPSGSDVSEQSSGGAHVFKPTV
RSEVTASSSGEDPPTYLSLSLPWTDVTRVNEPVQLNQNTVMDGGYTAE LFPVRKBEQVE
VEEEEEAKGISGGFGGFMTVVQEMIRTEVRSYMDLQRGNVGGSSSGGGGGSGCMPQSVN
SRRVGFREFIVNQIGIKME*

>G1842 (219..809)

ACTATTACATGCCTCTTCTCGCTTCAAAACGGCACCGTTTCCACTTGTATTATTTTTC
TCTCTATCGTCTAAACAAAAAAGCTGACTTGGGATTTTTTTTCAATTTGTCTAGCCCA
AAAGAAGAAGATAGAAACGAAGAAAAAAGCAAACACATTTTGGGTCCCCGGTGGTTAGG
ATCAAATTAGGGCACAACCTTATCGGAGAAAGAAGCCATGGGAAGAAGAAAAGTCGAGA
TCAAGCGAATCGAGAACAAAGCAGTCGACAAGTCATTTCTCCAAACGACGCAAAGGTC
TCATCGAAAAAGCTCGACAACCTTTCAATTCTCTGTGAATCTTCCATCGCTGTTGTGCGCC
TCTCCGGTTCCGGAAAACTCTACGACTCTGCCTCCGGTGACAACATGTCAAAGATCATTG
ATCGTTATGAAATACATCATGCTGATGAACCTTAAAGCCTTAGATCTTGAGAAAAAATTC
GGAATTATCTTCCACACAAGGAGTTACTAGAAATAGTCCAAAGCAAGCTTGAAGAATCAA
ATGTCGATAATGTAAAGTGTAGATTCTCTAATATCTATGGAGGAACAGCTCGAGACTGCTC
TGTCAGTAATTAGAGCTAAGAAGACAGAACTAATGATGGAGGATATGAAGTCACTTCAAG
AAAGGGAGAAGTTGCTGATAGAAGAGAACCAGATTCTGGCTAGCCAGGTGGGGAAGAAGA
CGTTTTCTGGTTATAGAAGGTGACAGAGGAATGTACGGGAAAAATGGCTCCGGCAACAAG
TACCGGAGACTCTTTTCGTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCACCAT
AAACTTACTCACAGCCTGATTGAGAGCTTTTACAAAATTGTAAATTATAAAAAGCTGCA
TAATAATCTCAACCTTTTTTATCTTCTCGCGCCAATGTGGAAATAAAGGTAAAAACAAAAC
GAAGCTCTTTTCTTTTATGCGAAAGAATGTAAAACTAAGATAAAGCTACCGATCTTTGT
TGTACCTTAGTAGACAAATATCAGAGTTCTTGTGCTTGT

>G1842 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRRKVEIKRIENKSSRQVTFSKRRKGLIEKARQLSILCESSIAVVAVSGSKLYDSASG
DNMSKIIDRYEIHHADELKALDLAEKIRNYLPHKELLEIVQSKLEESNVDNVSVDLSISM
EEQLETALSVIRAKKTELMMEDMKSLQEREKLLIENQILASQVGKKTFLVIEGDRGMSR
ENGSGNKVPETLSLLK*

>G1505 (1..681)

ATGGATGATATAGCGGAACCTTGAATGGTTATCAAATTTTCGTAGATGATTCTTCTTTCACG
CCGTATTCTGCTCCGACGAATAAACCGGTTTGGTTAACCGGAAATCGGAGACATCTTGTA
CAACCGGTTAAAGAGGAGACCTGCTTCAAATCCCAACATCCGGCCGTCAAACACGACCC
AAACGAGCCAGAACCGGAGTCAGAGTCTGGTCTCATGGTTCGAGTGGTTAACCGACTCA
TCTTCAAGCTCTACAACATCTTCGTGCTCCTCTCCTCGTCTTCAAGCCCTCTATGGCTC
GCCAGCGGTCACTTCTTGATGAGCCAATGACTAAAAACAAAAAGAAGAAGAAAGTTTGG

AAAAACGCTGGTCAGACGCAAACGCAAACGAGACGACGCGGCAGTGTGGTCATTGT
GGAGTTTCAGAAAACGCCGAGTGGAGAGCAGGACCATTAGGAGCGAAGACGTTGTGTAAT
GCGTGTGGTGTGCGTTACAAATCGGGTCGGTTACTACCCGAATATAGACCCGCTTGTAGC
CCAACATTTTCGAGTGAGCTTCACTCAAACCACCACAGTAAAGTCATTGAGATGCGTAGG
AAGAAAGAGACTTCTGACGGTGCTGAAGAAACCGGTTTGAACCAGCCGGTTCAGACGGTT
CAGGTTGTCTCGAGTTTTTTGA

>G1505 Amino Acid Sequence (domain in AA coordinates: TBD)
MDDIAELEWLSNFVDDSSFTPYSAPTNKPVWLTGNRRHLVQPVKEETCFKSQHPAVKTRP
KRARTGVRVWSHGSQLTDSSSSSSTSSSSSPRPSPLWLASGQFLDEPMTKTQKKKKVW
KNAGQTQTQTQTQRCGHGCVQKTPQWRAGPLGAKTLCNACGVRYKSGRLLPEYRPACS
PTFSSSELHSHNHSKVIEMRRKKETSDGAEETGLNQPVQTVQVVSFF*

>G657 (1..2331)

ATGAAGCGTGAGATGAAAGCACCTACTACTCCACTAGAGAGTCTCCAAGGTGACCTCAAA
GGAAAACAAGGGAGGACATCTGGCCCTGCTAGACGATCTACCAAAGGACAATGGACACCT
GAAGAGGACGAAGTCTTGTGTAAAGCTGTTGAGCGTTTTCAAGGAAAGAACTGGAAGAAG
ATAGCTGAATGTTTTTAAGGATCGGACTGATGTTTCAAGTGTCTTATAGATGGCAAAAGGTC
TTGAACCCAGAGCTTGTGAAAGGACCGTGGTCAAAAGAGGAGGATAACACAATAATTGAC
CTGGTTGAAAAATATGGGCCAAAGAAATGGTCTACTATATCTCAGCATTTTACCTGGGCGC
ATAGGAAAGCAATGTAGGGAAGGTGGCATAACCATCTTAACCTGGGATTAATAAAAAAT
GCATGGACTCAGGAAGAGGAACTGACTCTTATTCGTGCGCATCAAAATTTATGGGAATAAA
TGGGCAGAGCTTATGAAATTTTTGCCAGGAAGGTGAGATAATTTCGATAAAAAATCATTGG
AACAGCTCAGTTAAGAAGAAGTTGGATTCTACTATGCATCAGGTCTTTTAGATCAGTGT
CAAAGCTCGCCATTAAATGCGCCTTCAGAACAAATCTATCGCTTCATCTTCTCGTGGATG
CACAGCAATGGAGATGAAGGTAGTTCAAGGCCAGGGGTTGATGCTGAGGAATCAGAATGC
AGCCAAGCTTCAACTGTTTTCTCACAATCAACCAACGATTTACAAGATGAAGTTCAACGT
GGAAATGAGGAATATTACATGCCTGAATTTCAATCAGGAACGGAGCAGCAATCTCAAC
GCTGCATCTCATGCAGAACCGTACTACCTTCCCTTTAAAGATGTCAAATTTGTTGTCCCC
GAAATTTCTTGTGAAACAGAATGTTTCAAGAAGTTTCAAGATCTTAATTTGTTCTCACGAG
CTAAGAACTACCACAGCTACGGAGGATCAATTGCCGGGTGTATCTAATGATGTAAACAG
GACCGTGGTCTAGAGTTATGACCCATAACATGGACAACGGTGGAAAAAACCAAGCACTT
CAACAAGATTTTTCAAGTTTCAGTAAGATTAAGTGATCAACCTTTTTTGTCAAACCTCGGAC
ACAGATCCAGAAGCTCAAACCTTTGATCACGGATGAGGAGTGTGTAGGGTTCTTTTCCA
GATAACATGAAAGATAGCAGTACATCTTCTGGTGAGCAAGGTGCGGAATATGGTTGACCT
CAAAACGGCAAAGGATCTCTTTGTTCTCAGGCTGCAGAAACCCATGCTCATGAACTGGA
AAAGTTCCAGCTTTACCGTGGCATCTTCAAGTTCTGAGGGCCTGGCGGGTCATAATTGT
GTCCCTTTGTTGGATTTCAGACTTGAAGGACTCACTTTTACCCCGTAATGATTTCAACGCT
CCTATACAAGGTTGTGCGCTTTTGGAGCTACCGAATTAGAATGTAAGACTGATACAAAT
GACGGTTTCATCGATACTTACGGACATGTAACCTCCCATGGCAATGATGATAATGGTGGT
TTCCAGAACAAACAGGGGCTGTATATATTTCCCAAGGATTTCTTGAAGCTAGTACCTTTG
AATAGTTTTTCTTCTCTCTTAGAGTGAACAAGATTTATTTTCTATTGACGATAAGCCG
GCTGAAAAAGACAAAGGAGCTCTTTGTTATGAACCTCCACGTTTCCAAGTGCAGATATT
CCTTTCTTCAGCTGTGATCTTGTACCATCAAATAGTGACTTACCGCAAGAGTACAGTCCC
TTTGGTATCCGTGATGATGATTTCTTCAATGAATTGTACAACCTCCGTTAAGGTTATGG
GATTCACCGTGTACGATAGGAGCCCTGATGTCATGCTTAATGATACTGCCAAAAGTTTTT
AGTGGTGCACCATCCATCTTAAAGAAGCGGCATCGAGACTTGCTTTCACCTGTGCTTGAT
AGAAGAAAAGACAAAAGCTTAAAGGGCTGCGACTTCTCTCTTGGCTAATGATTTTTTCG
CGCTTAGATGTAATGCTTGATGAAGGAGATGATGTCATGACCTCTCGTCCGTCAGAGTCT
CCTGAAGATAAAAAATATATGTGCTTCCCTTCCATAGCCAGAGATAACAGAAATTTGTGCA
TCAGCTCGGTTATATCAAGAAATGATTCCGATAGATGAGGAACCAAGGAAACCTTAGAA
TCAGGTGGAGTGACTTCTATGCAAAATGAAATGGATGTAATGACGGTGGTGTCTCAGCT
AAAAATGTAAGTCCGCTTTGTCTTGCATATTATCTGGTATCAGTTATAA

>G657 Amino Acid Sequence (domain in AA coordinates: TBD)
MKREMKAPTPPLESLQGLDKGKQGRSTSGPARRSTKGQWTPPEDEVLCMAVERFQGNWKK
IAECFKDRITDVQCLHRWQKVLNPELVKGPWSKEEDNTIIDLVEKYGPKKWSTISQHLPGR
IGKQCRERWHNHLNPGINKNAWTQEEELTLIRAHQIYGNKWAELMKFLPGRSDNSIKNHW
NSSVKKKLDSEYASGLLDQCQSSPLIALQNKSIASSSSWMSNGDEGSSRPVDAEES
SQASTVFSQSTNDLQDEVQRGNEEYMPFHSQTEQQISNAASHAEPPYPSFKDVKIVVP

EISCETECSKKFQNLNCSHELRTTTATEDQLPGVSNDKQDRGLELLTHNMDNGGKNQAL
QQDFQSSVRLSDQPFLLNSDTPDPAQTLITDEECCRVLPDNDMKDSSTSSGEQGRNMVDP
QNGKGSLSQAAETHAHETGKVPALPWHPPSSSEGLAGHNCVPLLDSDLKDSLLPRNDSNA
PIQGCRFLFGATELECKTDTNDGFIDTYGHVTSNGNDNGGFPEQOGLSYIPKDSCLKVPL
NSFSSPSRVNKIYFPIDDKPAEKDKGALCYEPPRFPADIPFFSCDLVPSNSDLRQEYSP
FGIRQLMISSMNCCTPLRLWDSFCHDRSPDVMLNDTAKSFSGAPSILKKRHRDLSPLVD
RRKDKKLKRAATSSLANDFSRLDVMLDEGDDCMTSRPSESPEDKNICASPSIARDNRNCA
SARLYQEMIPIDEEPKETLESQGVVTSMQNGCNDGGASAKNVSPSLSLHIWYQL*

>G1959 (141..1028)

CGTCGACTGTCCATAAATCCGGAGCCTGACCCGACGTTTGACCCGGATCCGAAACTCCCA
CAATCTCCATACCACCCAAATTCATCTCCCTAAAGCTTCTCTCAC'TTCCCGGGAAAA
TCGGCGACCAAAATTGGAATGTACTCAGCGATTTCGTCGCTTCCACTCGATGGTGGAC
ACGTTGGTGGTGACTACCATGGACCTCTTGACGGAACCAATCTTCCCGGTGACGCTTGT
TGTTTAAACGACTGACCTAAACCTCGTCTCCGGTGGACAACCTGAGCTTCATGAGAGAT
TCGTTGACGCCGTACTCAGCTCGGTGGTCTGACAAAGCGACTCCCAAACTATTATGA
GAACAATGGGAGTGAAGGCTCTACTCTCTACCACCTCAAATCACATCTTCAGAAATCC
GCCTAGGGAGGCAAGCTGGCAAAGAATCAACTGAGAACTCTAAAGATGCTTCTGTGTAG
GGGAGAGTCAGGACACAGGTTCACTTCGACATCATCAATGAGAATGGCGCAGCAGGAGC
AGAACGAGGGTTACCAAGTCACCGAAGCTCTACGTGCTCAGATGGAAGTCCAAAGAAGAC
TACACGATCAATTGGAGGTGCAACGGAGGCTCCAGCTGAGGATAGAGGCACAAGGAAAT
ACCTGCAATCGATTCTTGAAAAAGCTTGCAAGGCCTTTGACGAGCAAGCTGCTACTTTG
CTGGACTTGAGGCTGCTAGGGAAGAGCTATCAGAGCTAGCCATCAAAGTCTCCAATAGCT
CTCAAGGAACATCAGTCCCGTACTTCGATGCAACAAAGATGATGATGATGCCATCGTTGT
CAGAGCTTGCACTAGCAATAGACAACAAAAACAACATCACAACCAACTGTTAGTAGAAA
GCTCTCTGACTTCCATCACACATGGGAGCTCTATATCTGCTGCATCAATGAAGAAGCGTC
AACGTGGAGACAATTTGGGCGTAGGGTATGAATCAGGCTGGATTATGCCTAGTAGACCA
TTGGATAAAGTTTAGGAGAGGGAAAAAGTTTATTATGGGAAAGGTAGAGATAAGATTTAA
CTGTTCTTTTACTTGGCTTTGAGGGCCTGCGGCCGCT

>G1959 Amino Acid Sequence (conserved domain in AA coordinates:46-97)

MYSAIRSLPLDGGHVGVDYHGPLDGTNLPDACLVLTTDPKRLRWTTTELHERFVDAVTQ
LGGPDKATPKTIMRTMGVKGLTLYHLKSHLQKFRLLGRQAGKESTENSKDASCVGESQDTG
SSSTSSMRMAQQEQNEGYQVTEALRAQMEVQRRLLHDQLEVQRRLLQLRIEAQGYLQSI
KACKAFDEQAATFAGLEAAREELSELAIKVSNSSQGTSPYFDTAKMMMPSLSELAVAI
DNKNITTNCSVESSLTSITHGSSISAASMKKRQRGDNLVGVYESGWIMPSSTIG*

>G2180 (1..1440)

ATGGCTCCTGTCTCGTTACCTCCAGGTTTCCGATTCCATCCAACAGACGAGGAACTAATT
ACTTACTATCTAAAAAGAAAGATCAACGGTCTAGAAATCGAACTTGAAGTTATCGCTGAA
GTTGATCTTTACAAGTGTGAGCCATGGGACTTACCAGGGAAGTCTTGTCTCCGAGCAAA
GACCAAGAATGGTACTTCTTCAGCCACGAGACCGGAAGTATCCCAACGGCTCAAGGACA
AACC GGGAACATAAGGCGGTTATTGGAAGGCTACAGGTAAGACCGCCGAGTTAGTTGG
AGAGACCGAGCCATAGGAACCAAGAAGACATTGGTTTACTACCGTGGGCGCGGCCACAT
GGCATAAGAACTGGTTGGGTCTATGCACGAATATCGACTTGATGAAACAGAATGTGAGCCT
TCTGCATACGGCATGCGAGACGCATATGCACTTTGTGCTGTGTTCAAAAAGATTGTTATT
GAAGCTAAGCCAAGAGATCAACATCGGTCAATGTCCACGCGATGTGCAATGTGAGTGGT
AATTGCTCATCGAGTTTGTGACTTGTTCGGATCTCGAAATCAGTTCAACTACTCATCAA
GTTCAAAACACATTCCAACCGCGATTGGCAACGAGCGATTAACTCCAACGCAATCAGC
AACGAGGATTGGTCACAATACTACGGTCTTCTTATAGACCGTTCCCTACTCCATATAAG
GTTAACACAGAGATGGAATGTTCAATGTTACAACACAATATATCTACCACCGTTGCGT
GTAGAGAACTCTGCGCTTTAGTGATTCCGATTTCTTCAGAGTATGACTCACAACAACGAC
CATGGCGTTTTGATGACTTTTACTTTGCTGCAAGTAACCTCAACCAACAATAATAGCGTT
GGTGATCAAGTGATCCAGTTGGCAATTATGATGAACAATTAATAACATCTAACCGTCAT
ATGAACCAGACTGGTTATATAAAAGAGCAGAAGATCAGATCGAGTTTGGATAATACTGAC
GAAGATCCAGGATTTTCATGGTAACAATACCAATGACAACATAGATATCGATGATTTTCTC
TCGTTTGATATATATAACGAGGACAACGTGAATCAAATAGAGATAATGAAGACGTGAAT
ACAAATGAAACCCCTGATTTCATCGGGATTGAGGTGGTTGAAGAAGAACTAGATTTAAC
AACC AAATGCTCATCTGACATATCAAACGACAAAGATTCTATATCACCAGTCGTACCT
TGTCACACGTTGAAAGTTACGCTCAATCCTATTAGTCACAATGTGGAAGAGAGAACATTG

TTCATTGAAGAGGACAAAGATTCTTGGTTACAAAGAGCTGAGAAGATCACGAAGACAAAA
CTAACACTTTTTAGTTTAAATGGCTCAGCAATACTACAAATGTCTTGCTATTTTTTCTGA
>G2180 Amino Acid Sequence (conserved domain in AA coordinates:7-156)
MAPVSLPPGFRFHPTDEELITYYLKRKINGLEIELEVIAEVDLYKCEPWLPGKSLPSK
DQEWYFFSPRDRKYPNGSRTNRATKGGYWKATGKDRRVSWRDRAIGTKKTLVYYRGRAPH
GIRTGWMHEYRLDETECEPSAYGMQDAYALCRVFKKIVIEAKPRDQHRSYVHAMSNVSG
NCSSSFDTCSDLEISSTTHQVQNTFQPRFGNERFNSNAISNEDWSQYYGSSYRPFPTPYK
VNTEIECSMLQHNIIYLPPLRVENSASFSDSDFFTSMTHNNDHGVFDDFTFAASNSNHNSV
GDQVIHVGNVDEQLITSNRHMNQTYIKEQKIRSSLDNTDEDPGFHGNNNTNDNIDIDDFL
SFDIYNEDNVNQIEDNEDVNTNETLDSSGFVEVEETRENNQMLISTYQTTKILYHQVVP
CHTLKVHVNPISHNVEERTLFI EEDKDSWLQRAEKITKTKLTLFSLMAQQYYKCLAIFF*
>G1817 (1..1308)
ATGAAGGACGCAGAGAAGCGAGAGGTGATTGCATCATCATCATTACAAAGAAAGAGAAAC
AGAGGAAGAAGACTAAGGAAAAAGAAAGAAAGAAACGAGAAGCGAGTACTAATGGTTCCA
TCATCATTACCAAACGACGTGCTAGAGGAGATCTTTTAAAGATTTCCGGTTAAAGCCCTA
ATCCGACTCAAGTCTCTCTCGAAACAATGGAGATCGACGATCGAATCTCGCAGTTTTGAA
GAGAGACACTTQIEDNEDVNTNETLDSSGFVEVEETRENNQMLISTYQTTKILYHQVVP
GGAGAAGAAGATCCCATAGAGGAACCGGGATTTCGTCCAGACACTGACATTGGTTTTAGG
TTATTCTGCTTGAATCGGCTTCTCTTCTATCCTTTACTCGTCTCAATTTCCCTCAAGGG
TTCTTCAACTGGATCTACATATCTGAAAGCTGTGATGGCCTTTTCTGCATCCATTCCCCA
AAATCACATTCCGTATATGTAGTGAATCCGGCTACACGGTGGCTCCGCCTACTTCTCTCG
GCAGGGTTTCAGATTTTGATCCACAAGTTTAAACCCCACTGAACGTGAGTGAATGTAGTG
ATGAAATCAATCTTTCATCTAGCATTCTGTAAGGCCACCGATTACAAATTAGTGTGGTTG
TACAAATTGTGATAAGTACGTTGATGATGCGTTCGAGTCCAAACGTGGGAGTCACAAAGTGC
GAGATTTTTGACTTTAGGAAAAATGCTTGGAGGTACTTGGCTTGCCTCCAAGTCATCAG
ATATTCTATTACCAAAGCCAGCATCTGCAAACGGGTCGGTTTATTGGTTTACAGAACCA
TATAATGAAAGAATCGAAGTAGTGGCTTTTGATATTCAGACCGAAACATTCCGGTTGCTG
CCTAAGATTAATCCGGCTATTGCTGGTTTCAGATCCTCACCATTGACATGTGCACTCTG
GATAATAGTTTGTGTATGTGCGAAAAGGAGAAAGATACTATGATCCAGATATTTGGAGG
TTGAAACCATCAGAAGACATGGGAAAAGATTTTTAGCATAGACTTGGTTTCTGTCTCT
TCTTCTCGGACTGAGAAGCGTGATCAATTTGATTGGAGCAAGAAGGATAGGGTTGAGCCA
GCCACACCCGTCGCGGTTTGTAAGAATAAGAAGATCCTTCTCTCACATCGCTATTCGGA
GGTTTGGTAAAGTACGATCCCCTAACAAAATCTATCGATTTTTTTTTCCGGACATCCTACC
GCTTACAGAAAAGTTATTTATTTTCAAAGTTTGATATCTCATCTATAA
>G1817 Amino Acid Sequence (conserved domain in AA coordinates:47-331)
MKDAEKREVIASSSLQQRNRGRRLRKRNRNEKRVLMVPSSLPNDVLEEIFLRFVKAL
IRLSLSKQWRSTIESRSFEERHLTIKAKAFVDHPKVMLVGEEDPIRGTGIRPDTDIGFR
LFCLESASLLSFTRLNFPQGFNWIIYISESCDGLFCIHSPKSHSVYVNPATRWLRLLPP
AGFQILIHKNFPTEREWNVVMKSIHFLAFVKATDYKLVWLYNCDKYIVDASSPNVGVTKC
EIFDPRKNAWRYLACTPSHQIFYQKPPASANGSVYWFTEPYNERIEVVAFDIQETFRLL
PKINPAIAGSDPHHIDMCTLDNSLCMSKREKDTMIQDIWRLKPSEDTWEKIFSIDLVSCP
SSRTEKRDQFDWSKKDRVEPATPVAVCNKKILLSHRYSRGLVKYDPLTKSIDFFSGHPT
AYRKVIYFQSLISHL*
>G1649 (61..1311)
ATTCACAAAAACCGGAAAAAAGACAAGTAAAGAAAGCTTTGTTCAAGTTTACTTCA
ATGGAAGCAAAACCTTAGCATCATCATCTGAACCAACATGATTTCTCCATCATCA
AACATTAACCAAAATTAAGAGATGAAGATTATATGGAGCTGGTGTGTGAAAATGGGCAG
ATTCTTGCAAAGATFCGAAGACCAAGAACAACGGTTCTTTTCAAAGCAACGTAGGCAA
TCTCTCCTGGATTTGTATGAGACCGAGTACAGCGAGGGTTTCAAGAAAAACATCAAGATT
CTTGAGACACACAAGTTGTTCCGGTGAGTCAGTCTAAGCCACAACAAGATAAAGAAACC
AATGAACAAATGAACAACAATAAGAAGAAGCTAAAGTCTCCAAAATCGAATTTGAGAGA
AATGTTTTCGAAAAGCAACAATGTGTGAATCATCAACATTAATTGATGTTTCTGCTAAA
GGTCCAAAGAATGTTGAAGTTACTACAGCTCCTCCTGATGAGCAATCTGCAGCTGTGGT
AGATCCACGGAATTGATTTTGTCTTCTCATCGAAGTTTTCTCGAGGAACCTCGAGAGAT
CTAAGTTGTTGTTCTTTAAAGAGGAAGTATGGAGATATTGAAGAAGAAGAATCAACCTAT
TTAAGTAATAATTGAGATGATGAATCAGATGATCGAAGACACAAGTTTCATGCGAGAACA
AGAAAGCCGGTGACTAAAAGAAAACGAAGCACAGAAGTCCATAAGTTATATGAAAGAAAA

CGAAGAGATGAATTCAACAAGAAAATGCGTGCTTTGCAGGACCTACTACCAAATTGTTAC
AAGGATGATAAGGCTTCATTGTTGGATGAGGCTATCAAATATATGCGGACCCTTCAACTT
CAAGTTCAGATGATGAGTATGGGAAATGGATTAATAAGACCACCTACGATGTTGCCAATG
GGTCATTACTCTCCCATGGGTCTAGGAATGCATATGGGTGCAGCAGCAACACCAACATCA
ATACCGCAATTCTGCCTATGAATGTTCAAGCAACCGGTTTTCCGGGGATGAACAATGCA
CCACCACAAATGCTAAGCTTTCTTAATCACCCAAGTGGACTAATCCAAACACTCCTATC
TTTTCTCCATTGGAAAATTGCTCTCAGCCATTCTGGTGCCTTCGTGTGTTTCTCAGACT
CAGGCTACTTCTTTTACTCAATTTCCCAAAGTCTGCGTCCGCCTCAAACCTAGAAGATGCA
ATGCAATATAGAGGAAGCAACGGTTTTAGTTATTATCGCTCGCCAACTAATGATTTGTA
GAAAGTTGATGTTTCTCCAACCTAACTTAAGCAAAAAAATGATCGTCTACTCT
GTGTGTTAGTCTATGGGCTTTTGGGCTTGATCTTGGAAACGATTTGAACCTAATTTCCA
ACTATTTTCAAAGTGGATGTACAAAGTAAAA

>G1649 Amino Acid Sequence (conserved domain in AA coordinates:225-295)

MEAKPLASSSEPNMISPSSNIKPKLKDEDYMLVCENGQILAKIRRPKNNGSFQKQRRQ
SLLDLYETEYSEGFKNILKILGDTQVVPVSQSKPQDKETNEQMNNKKKLKSSKIEFER
NVSKSNKCVESSTLIDVSAKGPKNVEVTTAPPDEQSAAVGRSTELYFASSSKFSRGTSRD
LSCCSLKRKYGDIEEEESTYLSNNSDDESDAKTQVHARTRKPVTKRKRSTEVHKLIERK
RRDEFNKKMRALQDLLPNCYKDDKASLLDEAIKYMRTLQLQVQMMSMGNGLIRPPTMLPM
GHYSPMGLGMHMGAAATPTSIPQFLPMNVQATGFPGMNAPPQMLSFLNHPGLIPNTPI
FSPLENCSQPFVVPSCVSQTQATSFTQFPKSASASNLEDAMQYRGSNGFSYYRSPN*

>G2131 (69..1010)

GTCTCTCATTTTTCATAATTCATTTCAGGATTGTCTCTCAATCTTTTATTCTTCTCATT
CACCGGTAATGGCAAAAGTCTCTGGGAGGAGCAAGAAAACAATCGTTGACGATGAAATCA
GCGATAAAACAGCGTCTGCGTCTGAGTCTGCGTCCATTGCCCTAACATCCAAACGCAAAC
GTAAGTCGCCCGCCTCGAAACGCTCCTCTTCAACGCAGCTCCCTTACAGAGGCGTCACAA
GGCATAGATGGACTGGGAGATACGAAGCGCATTTGTGGGATAAGAACAGCTGGAACGATA
CACAGACCAAGAAAGGACGTCAAGTTTATCTAGGGGCTTACGACGAAGAAGAAGCAGCAG
CACGTGCCCTACGACTTAGCAGCATTTGAAGTACTGGGACGAGACACACTCTTGAACCTCC
CTTTGCCGAGTTATGACGAAGACGTCAAAGAAATGGAAGGCCAATCCAAGGAAGAGTATA
TTGGATCATTGAGAAGAAAAAGTAGTGGATTTTCTCGCGGTGTATCAAAATACAGAGGCG
TTGCAAGGCATCACCATAATGGGAGATGGGAAGCTAGAATTGGAAGGGTGTGTTGGTAATA
AATATCTATATCTTGAACATACGCCACGCAAGAAGAAGCAGCAATCGCCTACGACATCG
CGGCAATAGAGTACCGTGGACTTAACGCCGTTACCAATTTTCGACGTGAGCCGTTATCTAA
ACCTTAACGCCCGCGCGGATAAAGCCGATTCCGATTCTAAGCCCATTCTGAAGCCCTAGTC
CGAGCGCCGAATCGTCCGATGATAACAAATCTCCGAATCAGAGGAAGTAATCGAACCAT
CTACATCGCCGGAAGTGATTCCAACCTCGCCGGAGCTTCCCCGACGATATCCAGACGTATT
TTGGGTGTCAAGATTCCGGCAAGTTAGCGACTGAGGAAGACGTAATATTCGATTGTTTCA
ATTCTTATATAAATCTGGCTTCTATAACGAGTTTGATTATGGACCTTAATCGTATTTTC
TACAAGTTTGTGTTTGATTATCTACACAATACATCAATATATTTCT

>G2131 Amino Acid Sequence (conserved domain in AA coordinates:50-186, 112-183)

MAKVSGRSKKTIVDDEISDKTASASESASIALTSKRKRKSPPRNAPLQRSSPYRGVTRHR
WTGRYEHLWLDKNSWNTQTKKGRQVYLGAYDEEEAAARAYDLAALKYWGRDILLNFPLP
SYDEDVKEMEGQSKEEYIGSLRRKSSGFSRGVSKYRGVARHHNGRWEARIGRVFGNKYL
YLGTYATQEEAAIAYDIAAIEYRGLNAVTFNFDVSRYLNPNAADKADSDSKPIRSPSREP
ESSDDNKSPKSEEVIEPSTSPVPIPTRRSFPDDIQTYFGCQDSGKLATEEDVIFDCFNSY
INPGFYNEFDYGP*

>G215 (1..1110)

ATGACTCGTCTGTTGCGATTGTAGCAACAATGGGCACAATTCACGCACGTGTCCAACG
CGTGGGTCTGGTTCTCTCCTCCGCCGTGAAGTTATTGGTGTGAGGTTAACGGATGGCTCG
ATTATTAAGAAAGAGTGCAGATATGGGTAATCTCTCGGCATTGCGTGTGCGGCGGCGCG
GCAACGCCACCACGTTTATCTCCGTCGTCTCTCTGGCGACGTCAAATCTTAATGATTGCG
CCGTTATCGGATCATGCCCCGATACTCTAATTTGCATCATAATGAAGGGTATTTATCTGAT
GATCTCTGCTCATGGTTCTGGGTCTAGTCACCGTCGTGGTGAGAGGAAGAGAGGTGTTCTCT
TGGACTGAAGAGGAACATAGACTATTCTTAGTCGGTCTTCAGAACTCGGGAAAGGAGAT
TGGCGCGGTATTTGAGAAACTATGTAACGTCAAGAACTCTACACAAGTGGCTAGTCAT
GCTCAAAAGTATTTTATTCGACATACTAGTTCAAGCCGCAGGAAAAGACGGTCTAGCCTC
TTCGACATGGTTACAGATGAGATGGTAACCGATTATCGCCAACACAGGAAGAGCAGACC

TTAAACGGTTCCTCTCCAAGCAAGGAACCTGAAAAGAAAAGCTACCTTCCTTCACTTGAG
CTCTCACTCAATAATACCACAGAAGCTGAAGAGGTCGTAGCCACGGCGCCACGACAGGAA
AAATCTCAAGAAGCTATAGAACCATCAAATGGTGTTCACCAATGCTAGTCCCGGGTGGC
TTCTTTCTCCTTGTTCCTCAGTGACTTACACGATTTGGCTCCCTGCGTCACTTCACGGA
ACAGAACATGCCTTAAACGCTGAGACTTCTTCTCAGCAGCATCAGGTCCTAAAACAAAA
CCTGGATTGCTAAAGAACGTGTGAACATGGACGAGTTGGTCGGTATGTCTCAGCTTAGC
ATAGGAATGGCGACAAGACACGAAACCGAAACTTCCCCCTCCCCGCTATCTTTGAGACTA
GAGCCCTCAAGCCATCAGCGTTTCACTCGAATGGCTCGGTTAATGGTGCAGATTTGAGT
AAAGGCAACAGCGCGATTTCAGGCTATCTAA
>G215 Amino Acid Sequence (domain in AA coordinates: TBD)
MTRRCSHCSNNGHNSRTCPTRGSGSSSAVKLFGVRLTDGSIKKASMGNLSALAVAAAA
ATHHRLSPSSPLATSNLNDSPSLDHARYSNLHHNEGYLSDDPAHSGSGSSHRRGERKRGVP
WTEEHRLFLVLGLQKLKGKDWRGISRNYVTSTPTQVASHAQKYFIRHTSSSRKRKRRSSL
FDMVTDEMVIDSSPTQEEQTLNGSSPSKEPEKSYLPSLELSLNNTEAEVVATAPRQE
KSQEAIEPSNGVSPMLVPGGFFPPCFPVYTIWLPASLHGTEHALNAETSSQHQVLKPK
PGFAKERVNMDELVGMSQLSIGMATRHETETSPSPSLRLEPSRPSAFHSNGSVNGADLS
KGNSAIQAI*
>G1508 (1..420)
ATGCTAGATCACAGTGAAGAGGCTCTTATTGGTTGATTGAGAAACCATGAAAACAAGAGCT
GAAGATATGATCGAACAGAACACACTAGTGTAAACGACAAGAAGAGACTTGTGCTGAT
TGTGGAACCAAGTAAACTCCTCTTTGGCGTGGTGGTCTGTTGGTCCAAAGTCGTGTGT
AACGCGTGTGGGATCAGAAACAGAAAGAGAAGAGGAGGAACAGAGATAATAAGAAA
TTAAAGAAATCGAGTTCTGGCGGCGAAACCGTAAATTTGGTGAATCGTTAAACAGAGT
TTGATGGATTTGGGGATAAGGAAGAGATCAACGGTGGAGAAGCAACGACAGAGCTTGGT
GAAGAGAACAAGCCGTGTGTTACTCATGGCTCTTTCTTATGGCTCTGTTTACGCTTAG
>G1508 Amino Acid Sequence (domain in AA coordinates: 38-63)
MLDHSEKVLVLDSETMKTRAEDMIEQNNTSVNDKKKTCADCGTSKTPLRGGPVGPKSLC
NACGIRNRKRRGGTEDNKKLKKSSSGGGRKFGESLKQSLMDLGIRKRSTVEKQRQKLK
EEEQAVALLMALSYGSVYA*
>G2110 (36..1622)
GAGAGCTAATAAAAAATTTATCAAGAAGACTAATATGGAGAAGGACGATTTCTTGAGGA
GTGGTCATGGAAGAGAAGAAAGCCATGATGAGATGAGAAAACCTGATTCTCTCACGATG
ATTCTCATCAAGAACACGACCATATTATAAGATCCAAGTTGGACTCAACTAAAGTCGAAA
TGGATGAGGCTAAAGAGGAAAAATCGAAGACTAAAGTCATCATTGAGTAAATCAAGAAAG
ATTTTGACATCCTTCAAACACAATACAACCAATTAATGGCCAAACATAACGAACCAACCA
AGTTCCAATCAAAAAGGGCATCATCAAGACAAAGGCGAAGATGAAGACAGAGAAAAAGTTA
ACGAACGTGAAGAAGCTTGTCTCGTTGAGCCTAGGCGAGACGGTTAAATTCAGAGGTTCCAA
GTGGTTCGAATAAAGAAGAAAAAATAAGATGTTGAAGAAGCGGAAGGTGACAGAAATT
ATGATGATAATGAAAAAAGCAGTATTCAAGGGTTGAGTATGGGGATTGAATACAAGGCTT
TGAGTAATCCTAATGAGAAGTTAGAGATTGATCATAATCAAGAAACCATGTCGTTGGAGA
TTAGTAACAATAAAGATCAGATCAGAAATAGTTTTGGGTTTAAAGAAATGATGGAGATG
ATCATGAAGATGAAGATGAGATTTTGCCCTCAAAACCTTGTTAAGAAAACCTAGGGTTTCGG
TGAGATCAAGATGTGAGACACCAACGATGAACGACGGATGTCAATGGAGGAAATATGGCC
AGAAAATAGCTAAAGGCAATCCATGTCCCCGAGCTTACTATCGTTGCACCATTGCAGCTT
CTTGTCAGTAAGAAAACAGGTGCAAAGATGTTCAGAAGATATGTCTATACTTATCTCAA
CGTACGAAGGAACACATAACCATCCACTTCCCATGTGACCAACTGCCATGGCCTCTGCCA
CTTCCGCTGCCGCTCCATGCTTCTCTCCGCGCCTCTCCTCTCATCCGCCGCGAGCTG
ATCTTCATGGCCTTAACCTTCTCTTTCCGCAACAACATCACTCCAAAACCTAAACTC
ATTTCTCCAATCCCCCTTCTTCTTCTGGCCATCCGACCGTCACTCTCGACCTCACAACT
CCTCCTCGTCGACGAACCGTTCTTATCAATGCTCAATAGATTGAGCTCTCCTCCAAGTA
ATGTCTCAGCATCTAATAGTTATCTTCAACCAATCTCAACTTTTCAAACAACCAACA
CATTGATGAATTGGGGTGGTGGTGAATCCCAAGTGATCAATACCGTGCAGCTTACGGCA
ACATTACACCCATCAGCAATCACCTTACCACAAAATCATTCAAACCCGAACCGCGGGT
CATCTTTCGATCCGTTTGAAGATCATCTTCATCACATTCTCCACAAATAAATCTTGATC
ATATCGGAATCAAGACATCATCAGTCACCAAGTGCATCTTTACCGGCTGAAACAATCA
AGGCAATCAGCAGATCCAAGTTTCCAATCGGCTTTGGCGACAGCTCTATCTTCCATCA
TGGGCGGCGATTTAAAGATTGATCACAATGTGACTAGAAATGAAGCTGAGAAGACCCCTT

AAAGAGAATTGTTATATATATGTTCTTATATACTCAGTACATTGGTAAATGGGTTTAGAC
TTTCACTAGTTTCCTAGTTCATCTATATATTGGTTGTTTAAATCACAAGTTTATTTTGTG
TTGGAGTTTATGGAATAATGTGTACATATGAACTTTAGAACGAATAAATAAACTTGG
AATTCCTTTTTAAAAAAAAAAAAAAAAA

>G2110 Amino Acid Sequence (conserved domain in AA coordinates: 239-298)

MEKDDFLRSGHGREESHDEMRKLDSSHDDSHQEHDI IRSKLDSTKVEMDEAKEENRRLK
SSLKIKKDFDILQTOYNQLMAKHNEPTKFQSKGHHQDKGEDEDREKVNEREELVSLSLG
RRLNSEVPSGSNKEEKNKDVVEAEGRNYDDNEKSSIQGLSMGIEYKALSNPNEKLEIDH
NQETMSLEISNNKIRSONSFQFKNDGDDHEDEDEILPQNLVKKTRVSVRSRCETPTMND
GCQWRKYGQKIAKGNPCPRAYYRCTIAASCPVRKQVQRCSIEDMSILISTYEGTHNHPLPM
SATAMASATSAASMLLSGASSSSAAADLHGLNFSLSGNITPKPKTHFLQSPSSSGHP
TVTLDLTTSSSSSQPFSLMLNRFSSPPSNVSRNSYPTNLNFSNNTNTLMNWGGGNPS
DQYRAAYGNINTHQQSPYHKIIQTRTAGSSFPFGRSSSSHSPQINLDHIGIKNIISHQV
PSLPAETIKAITTDPFQALATALSSIMGGDLKIDHNVTREAEKSP*

>G2442 (71..997)

TCGACCAATTTAGACCATTCCAAATTCGTCGTCCTTTTCTCTGTGTAGTCTAATTATATA
TTACAAGTAGATGAATTGGTTACCTGAAGCTGAAGCTGAGGAGCACCTTGAAGGTATTCT
CTCTGGTGATTTCTTTGATGGTCTCACCACCTTGAAGGATGCGGACCTTGAAGACATCGA
TTCCACCAATGGTGAGGGAGATGGGTCGCCAGGTTTCAAGACCTTGAGCCTCCTCCCTT
GGATATGTTCCCTGCTTTGCCTTCTGACCTCACCTCTTGTCCCAAGGCGCCGCTCGTGT
GCGGATTCACCAACAACATGATTCCTGCTTTGAAGCAGTCTGTTCTTCTGAAGCCTTGT
CGGCATTATAGCACTCCCAACCAATCTTCAGCTCCTCCTGATATCAAAGTTTCATATCT
ATTTCACTCTAATCCAGTGTGAGTTCTCGAGAAGCTTATGGTTCTCTCTCCACCCA
AACTCCGGATCTCAGAGATTGGCTTTCCCTGTGAAGGCATGAGAAGCAAGCGCAGACG
CCCCAACAGCTGAGACTTAGCTACCTTTTCCCTTTGAACCCAGAAAGTCAACTCCGGG
TGAATCAGTAACCGAGGTTACTATTCTTCTGAGCAACATGCCAAGAAGAAGCGCAAGAT
TCATCTGATCAACACACCGAGTCTTCCACTTTGGAGTCAAGTAAGTCCGATGGGATAGT
CCGGATATGCACTCATTGTGAGACAATCACGACCCACAGTGGAGGCAAGGACCCAGTGG
ACCCAAGACCTCTGCAACGCTTGGCGAGTCCGGTTCAAATCTGGTCGCCTAGTTCCAGA
ATACCGGCCAGCCTCAAGCCCGACCTTCATCCCATCTGTGCATTCAAACCTCACACAGGAA
GATCATTGAGATGAGAAAGAAGGACGACGAGTTTGATACCAGCATGATTGCGAGTGATAT
CCAGAAGGTAAAGCAGGGGAGGAAGAAATGGTATAAAAGTA

>G2442 Amino Acid Sequence (domain in aa coordinates: 220-246)

MNWLPEAEAEHLKGLSGDFFDGLTNHLDCPLEDIDSTNGEGDWVARFQDLEPPPLDMF
PALPSDLTSCPKGAARVRIIPNNMIPALKQSCSSEALSGINSTPHQSSAPPDIKVSYLFQS
LTPVSVLENSYGLSTQNSGSQRLAPFVKGMRKRRRPTTVRLSYLFPFEPKSTPGESV
TEGYYSSEQHAKKKRIHLIHTTESSTLESSKSDGIVRICTHCETITTPQWRQGPSGPKT
LCNACGVRFKSGRLVPEYRPASSPTFIPSVHSNSHRKIIEMRKKODEFDTSMIRSDIQKV
KQGRKKMV*

>G1051 (66..1031)

CCTGTAAATTCAGATTTGCTTTCTTTGGTAATCTTTGGATCAAGATCCATCTATTTTTT
CTTCAATGGCACAACTCCCTCCTAAAAATCCCAACATGACACAACATTGGCTGATTCT
CTTCCAAAAGCTCTCTCTTTCTTACCCCAACCGCAACCGCTGTGCGCCACCGCTACAA
CCACCGTACAAAACCCCTCATGGGTGACGAATTCCTCGACTTCTCAGCGTCTCGCCGTG
GCAACCACCGTCCGTCCATCAGCGACTCTATCGCATTCTCGAAGCTCCAACAGTCAGCA
TCGAAGACCACCAATTGACAGGTTGATGACGAACAGTTTCATGTGATGTTACCGGACG
ACGACAACCTTCATAGCAATCTTCCCATATCAACAACAAAAATAACAATGTGGGGCCCA
CGGGATCTTCTCGAACACATCCACGCCGTCCAATAGCTTCAACGACGATAACAAAGAAT
TACCACCGTCCGATCATAACATGAACAATAATATCAACAACAACTATAACGATGAAGTCC
AAAGCCAATGCAAGATGGAGCCAGAAGATGGTACGGCGTGAATAACAATTCGGGTGATA
GCTCCGGCAACCGGATTCTCGATCCCAAAAGGTTAAGAGAATATTAGCAAAATCGGCAAT
CAGCACAGAGATCAAGGTGAGGAACTGCAATACATATCAGAGCTCGAACGTAGCGTCA
CTTCGTTGACGGCGGAAGTGTGAGTGTATCGCCAAGAGTTGCATTCTTGGATCATCAAC
GTTTGCTTTTAAAGCTTGACAACAGCGCTCTCAAGCAACGAATCGTGTCTTATCTCAAG
ACAAGCTTTTCAAAGACGCACATCAAGAAGCATGAAGAGAGAAATAGAGAGACTTCGAC
AAGTGTATAATCAACAAGCCTCACGAATGTGGAATGCAAAATCATTTATCGGCGACCG
GAGCCGGTGCTACTCCGGCCGTGACATCAAGTCGTCCGTTGAAACAGAGCAGCTCCTCA

ATGTCTCATAAATTAACCATCATGCATCATCAACATTTCTCTCTTTAGCTTCTTGG
CAAAAGTTCTTGACTATAAAATCTCTTTCCGGTAAGAAATTCAGGAGATATAACATTTTTT
ATTCTAATCACATTGTTTTTAAGTTGTGATGAATTCAGTTTGATGTATCTTATTTATTTT
GTTTATGTCGTCTTTTTTTCTTGGGGTTGATGGAAGGAATCATCAATTGTTGTTGTAC
AAAGAACTAGTTGAATTTTTTTTTTTTTTTT

>G1051 Amino Acid Sequence (domain in AA coordinates 189-250)
MAQLPPKIPNMTQHWPDFSSQKLSFPSTPTATAVATATTTVQNPSWVDFLDFSASRRGN
HRRSISDSIAFLCAPTVSIEDHQFDRFDEQFMSMFTDDNLHNSPNSHNNKNNNVGPTG
SSSNTSTPSNSFNDDNKELPPSDHNMNNNNNNYNDEVQSQCMEPEDGTASNNNSGDSS
GNRILDPKRVKRILANRQSAQRSRVRKLQYISELERSVTSLQAEVSVLSPRVAFLDHQRL
LLNVDNSALKQRIAAALSQDKLFKDAHQEALKREIERLRQVYNQQLTNVENANHLSATGA
GATPAVDIKSSVETEQLLNVS*

>G1052 (138..1127)

TGATCATCTAAACTTTCAATTTCTCTCTTGATCCTCACTTGAATTTTTTGTGTTTCTC
TCAAATCTTTGATCCTTTCTCTTTGTTTTTCATTTGACCTCTTACAAAAAATCTGGTGTG
CCATTAAATCTTTATTAATGGCACAACCTTCCTCCGAAAATCCCAACCATGACGACGCCAA
ATTGGCCTGACTTCTCTCCAGAACTCCCTTCCATAGCCGCAACGGCGGCAGCCGCAG
CAACCGTGGACCTCAACAACAAAACCTTCATGGATGGATGAGTTTCTCGACTTCTCAG
CGACTCGCGTGGGACTCACCGTCGTTCTATAAGCGACTCCATTGCTTTCTTGAACCAC
CTTCTCCGGCGTCGGAAACCACTTCGATAGGTTTGACGACGAGCAATTCATGTCCA
TGTTCAACGACGACGTACACAACAATAACCACAATCATCATCATCACAGCATCAACG
GCAATGTGGGTCCACGCGTTCATCTCCAACACCTCCACGCCGTCGGATCATAATAGCC
TTAGCGACGACGACAACAACAAGAAGCACCACCGTCCGATCATGATCATCACATGGACA
ATAATGTAGCCAATCAAAACAACGCCGCCGGTAACAATTACAACGAATCAGACGAGGTCC
AAAGCCAGTGCAAGACGAGCCACAAGATGGTCCGTCGGCGAATCAAAACTCCGGTGGAA
GCTCCGTAATCGTATTCACGACCCTAAAAGGGTAAAAGAATTTAGCAAATAGGCAAT
CAGCAGAGATCAAGGGTGAGGAAATTGCAATACATATCAGAGCTTGAAAGGAGCGTTA
CTTCATTGCAGACTGAAGTGTGAGTTATCGCCAAGAGTTGCGTTTTTGGATCATCAGC
GATTGCTTCTCAACGTCGACATAGTGCTATCAAGCAACGAATCGCAGCTTTAGCACAAG
ATAAGATTTTCAAAGACGCTCATCAAGAAGCATTGAAGAGAGAAATAGAGAGACTTCGAC
AAGTATATCATCAACAAAGCCTCAAGAAGATGGAGAATAATGTCTCCGATCAATCTCCGG
CCGATATCAACCGTCCGTTGAGAAGGAACAGCTCCTCAATGTCTAAAGCTGTTCTGTTCA
CTAAGATCTTTCTTTTCATGGCGAAAAGATTCTTGACTATAAAACCTCTTTGTGTCAAGA
AATTAATTTATCAAGAAGATGGCCTTTTTTATTTGATCTAATCACATTTTTTTAAGTTG
TGATGAATTTGCTTTTGATGTATCTGTTTTTTTTTTTTTTTTTTT

>G1052 Amino Acid Sequence (domain in AA coordinates 201-261)
MAQLPPKIPTMTTPNWPDFSSQKLPSIAATAAAAATAGPQQNPSWVDFLDFSATRRGT
HRRSISDSIAFLPEPPSSGVGNHHRFDRFDEQFMSMFDNDVHNNNNHHHHHSINGNVGPT
RSSSTSTPSDHNLSDDDNKEAPPSDHDHMDNNVANQNNAAAGNNYNESDEVQSQCKT
EPQDGPANQNSGSSGNRIHDPKRVKRILANRQSAQRSRVRKLQYISELERSVTSLQTE
VSVLSPRVAFLDHQRLLLNVDNSAIKQRIAAALQDKIFKDAHQEALKREIERLRQVYHQ
SLKKMENNVSQSPADIKPSVEKEQLLNVS*

>G1079 (1..1995)

ATGGGTTGTGCTGCTTCAAGAATTGATAATGAAGAAAAGGTTTATAGTGTGTAGGCAGAGA
AAGAGGCTAATGAAAAAGTTATTAGGGTTCAGGGGAGAATTGACAGATGCACAGTTGGCT
TATCTTAGAGCTTTGAGGAACACTGGTGTACTCTTAGGCAATTCAGTGTCTGAGACC
TTGGAGCTTGAAAAACTAGTTATGGTTTAAAGTTTGCCTTTGCCTCCTTCGCCTCCTCCT
ACATTGCCTCCTTCACCTCCACCACCTCCTCCATTTAGCCCGGATTTGAGAAATCCTGAG
ACTAGTCATGACTTGGCTGATGAGGAGGAAGAGGGTGAAAATGATGGTGGTAATGATGGA
AGTGGTGCAGCTCCTECGCCTCCATTGCCGAATCTTGGAACATTTGGAACCCCTTTGAG
TCACTTGAGCTGCATAGTCATCCAAATGGTGACAATGTAGTTACACAAGTTGAACTGAAG
AAGAAACAACAATTCAGCAAGCTGAAGAGGAAGATTGGGCGGAGACGAAGTCTCAATTT
GAGGAAGAAGATGAGCAACAAGAAGCAGGAGGTACTTGCCTTGATTTGAGTGTTCATCAA
ATAGAGGCTGTTAGTGGCTGTAAATGAAGAAGCCACGTCGTCTGAAGTTAAGCTGGGA
GAAGTTATGGACGGTAACCTCATCTATGACAAGCTGCTCCGGTAAAGATCTTGAGAAAAC
CATGTGACTGATTGTAGAATCAGGAGGACCTTAGAAGGAATCATCAGAGAGTTGGATGAT
TATTTTCTTAAAGCATCGGGTTGCCGAGAAGGAGATAGCTGTGATAGTAGACATCAACAGT

AGGGATACTGTTGATCCTTTTCAGGTACCAGGAAACAAGAAGGAAGAGAAGCAGCTCGGCA
AAGGTATTCACTGTCATTGTTCATGGAGTTGGTCTTCAAAGTCTCTTCAGTTGGGCAAAGAT
GCTACAACAAGCGGGACTGTTGAACCCCTGTAGGCCTGGAGCTCACTGCAGCACACTTGAG
AAGCTATACACAGCTGAGAAGAACTTTACCAGCTAGTCAGAAACAAGAGATTGCCAAA
GTGGAGCATGAGAGGAAGTCTGCATTACTGCAAAAGCAAGATGGGGAAACCTATGATTTG
AGCAAAATGGAGAAAGCACGCTTGTCTTTTGAGAGTTTGGAAACCGAGATACAGCGTCTA
GAAGATCCATAACTACAACAGCTCATGTTTGCTTAACCTTGATCAATGATGAGCTGTAT
CCGCAGCTAGTTGCTTTAACTTCAGGGCTAGCACAGATGTGGAAAACAATGCTCAAGTGT
CATCAAGTTCAAATTTCATATATCCCAGCAACTGAACCATCTTCCGATTACCCGAGTATA
GATCTCAGTTCGGAATACAAACGCCAGGCGGTTAATGAACTAGAGACCGAGGTTACTTGC
TGGTACAATAGCTTTTGCAAGTTAGTAAATTCACGAGGAGAATACGTGAAAACACTCTGT
ACGTGGATCCAACCTTACTGATCGCCTCTTAACGAAGACAACCAAAGAAGTAGCTTGCCT
GTTGCTGCTCGTAAGCTCTGCAAAGAGTGGCAGCTTGAATACAACCTGCGTAGGAAATGC
AATAAAGTTGAGAGGAGGCTTGAGAAAGAGCTAATTTCACTGGCTGAGATTGAAAGAAGG
CTCGAGGGGATTTTAGCAATGGAAGAGGAGGAAGTAACTCAACGAGTTTGGGCTCTAAG
CATCCGTGTCAATCAAACAAGCCAAGATCGAAGCCTTGAGAAAACGAGTGGATATTGAG
AAAAGTAACTTAACTCGGTCGAGGTTAGTAAGAGAATGACACTAGACAACCTCAA
TCAAGCCTTCCCAATGTCTTTCAGATGTTGACTGCTCTAGCTAATGTCTTTGCCAATGGG
TTGAAATCCGTTAATGGCCAAACCGGTACAGATGTTTCCGACACATCCCAACATTCCGAT
GAATCTCAACCCTAA

>G1079 Amino Acid Sequence (conserved domain in AA coordinates:1-50)

MGCAASRIDNEEKVLVCRQRKRLMKLLGFRGEFADAQLAYLRALRNTGVTLRQFTESET
LELENTSYGLSLPLPPSPPTLPSPPPPPFSPDLRNPETSHDLADEEEEEEGENDGGNDG
SGAAPPPLPNSWNIWNPFFESLELHSHPNGDNVVTQVELKKKQIQQAEEEDWAETKSQF
EEDEQQEAGGTCLDLSVHQIEAVSGCNMCKPRRLKFKLGEVMDGNSSMTSCSGKDLEKT
HVTDCRIRRTLEGI IRELDDYFLKASGCEKEIAVIDINSRDTVDPFRYQETRRKRSSA
KVFSALSWSWSSKSLQLGKDATTSGTVEPCRPGAHCSTLEKLYTAEKKLYQLVRNKEIAK
VEHERKSALLQKQDGETYDLSKMEKARLSLESLETEIQRLSDSITTRSCLLNLINELEY
PQLVALTSGLAQMWKMLKCHQVQIHIISQQLNHLDPYPSIDLSSEYKQAVNELETEVTC
WYNSFKLVNSQREYVKTLCTWIQLTDRLSNEDNQRSSLPVAARKLCKEWQLEYNLRRKC
NKLERRLEKELISLAEIERRLEGILAMEEEVSSLSLGSKHPLSIKQAKIEALRKRV DIE
KTKYLSNVEVSKRMTLDNLKSSLPNVFQMLTALANVFANGFESVNGQTGTDVSDTSQHSD
ESQP*

>G1335 (56..667)

TTTTTTTTTAAAAGATTTAGAGAGAAAAGTGAGTTATTAAGAGATTCCAATCAAATGAG
CGGAGACAACGGCGGTGGTGAGAGGCGCAAAGGCTCCGTCAAGTGGTTTGATACCCAGAA
GGGTTTCGGCTTCATCACTCTGACGACGGTGGCGACGATCTCTTCGTTCAACAGTCCTC
CATCAGATCTGAGGGTTTCCGTAGCCTCGCTGCCGAAGAAGCCGTAGAGTTCGAGGTTGA
GATCGACAACAACAACCGTCCCAAGGCCATCGATGTTTCTGGACCCGACGCGCTCCCGT
CCAAGGAAACAGCGGTGGTGGTTTCATCTGGCGGACGCGCGGTTTTCGGTGGAGGAAGAGG
AGGTGGACGCGGATCTGGAGGTGGATACGGCGGTGGCGGTGGTGGATACGAGGAAGAGG
AGGTGGTGGTTCGAGGAGGCGGACTGCTACAAGTGTGGTGGAGCCCGTCACATGGCGAG
AGACTGTTCTGAAGGCGGTGGAGGTTACGGAGGAGGCGCGGTGGCTACGGAGGTGGAGG
CGGATACGGCGGAGGAGGTGGTGGTTACGGAGGTGGTGGCGGTGGAGGTGGTGGCGCGG
GGGAAGCTGCTACAGCTGTGGCGAGTCCGGACATTTCCGCCAGGGATTGCACCAGCGGTGG
ACGTTAAAACCAACCGCGGTACGCGGTGGAGAAGAGTGAGTTGGTTATCTCACAAGTGA
TCGGTTCTTTCTCCCGCCGCTTCTATCTCTCTATTATCCACTTTTGTCTATTATGATG
GATCTCTATCTTTGTTAGTTGGTTTTTTCTTGATGGTTTCGGATTAGGACTCTTCTTTTG
GTTTTGCTACTTATGGTTGGTTTTATTTATGGTACTTGTGATATGGGTGAAATGCTCTAC
TTGTTGCTCTGTTTCAAGTGTTCATAATATGCCAACAATATTCTGGGTTTTGTTTCAA
AAAAA

>G1335 Amino Acid Sequence (domain in AA coordinates: 24-43, 131-144, 185-203)

MSGDNNGGERRKGSVKWFDTKGFGFITPDDGGDDLFWHQSSIRSEGFRLAAEEAVEFE
VEIDNNRPFKAIDVSGPDGAPVQNSGGSSGGRGFGGGRGGRGSGGGYGGGGGGYGG
RGGGGRGSDCYKCGEPGHMARDCEGGGGYGGGGGGYGGGGYGGGGGGYGGGGGGGG
GGGSCYSCGESGHFARDCTSGGR*

>G157 (31..621)

GGGCATAACCCCTTATCGGAGATTTGAAGCCATGGGAAGAAGAAAAATCGAGATCAAGCGA
ATCGAGAACAAAAGCAGTCGACAAGTCACTTTCTCCAAACGACGCAATGGTCTCATCGAC
AAAGCTCGACAACCTTCGATTCTCTGTGAATCCTCCGTCGCTGTTGTCGTATCTGCC
TCCGGAAAACTCTATGACTCTTCCCTCCGGTGACGACATTTCCAAGATCATTGATCGTTAT
GAAATACAACATGCTGATGAACCTTAGAGCCTTAGATCTTGAAGAAAAAATTCAGAATTAT
CTTCCACACAAGGAGTTACTAGAAACAGTCCAAAGCAAGCTTGAAGAACCAAATGTCGAT
AATGTAAGTGTAGATTCTCTAATTTCTCTGGAGGAACAACCTTGAGACTGCTCTGTCCGTA
AGTAGAGCTAGGAAGGCAGAACTGATGATGGAGTATATCGAGTCCCTTAAAGAAAAGGAG
AAATTGCTGAGAGAAGAGAACCAGGTTCTGGCTAGCCAGATGGGAAAGAATACGTTGCTG
GCAACAGATGATGAGAGAGGAATGTTCCGGGAAGTAGCTCCGGCAACAAAATACCGGAG
ACTCTCCCGCTGCTCAATTAGCCACCATCATCAACGGCTGAGTTTACCTTAAACTCAA
AGCCTGATTCAATTAAGAGAATAAATTTGTATATTATAAAAAGCTGTGTAATCTCAAA
CCTTTTATCTTCTCTAGTGTGGAATTTAAGGTCAAAAAGAAAACGAGAAAGTATGGATC
AGTGTGTACCTCTCTCGGAGACAAGATCAGAGTTTGTGTGTTGTGTCTGAATGTACGG
ATTGGATTTTTAAAGTTGTGCTTTCTTTCTTCAAAAAA

>G157 Amino Acid Sequence (domain in AA coordinates: 2-57)
MGRRKIEIKRIENKSSRQVTFKRRNGLIDKARQLSILCESSVAVVVVSASGLYDSSSG
DDISKIIDRYETIQHADELRLDLEEKIQNYLPHKELLETVQSKLEEPNVDNVSDSLISL
EEQLETALSVSRARKAELMMEYIESLKEKEKLLREENQVLASQMGKNTLLATDDDERGMFP
GSSSGNKIPETLPLLN*

>G1895 (1..954)

ATGAATAACCAATCTGTTACTGACAATACAAGTCTTAAGCTGTCATCTAATCTTAACAAC
GAGTCAAAAGAAACATCTGAGAACAGTGTATGACCAACACAGCGAGATCACAAATTACA
TCGGAAGAAGAGAAAAACAACCTGAAGTGAAGAAACAGACAAGATTCTTCCATGTCCGAGA
TGCAACAGCGCAGACACCAATTTCTGTTACTACAACAACCTACAACGTTAACCAGCCACGT
CACTTCTGTAGAAAATGCCAGAGGTATTGGACCGCTGGTGGATCCATGAGGATCGTCCCG
GTTGGCTCAGGCCGTCGCAAGAACAAGGGATGGGTTTCTTCAGACCAGTACCTGCACATC
ACTTCCGAGGATACTGACAATTACAATAGCTCCTCAACAAAGATTCTAAGCTTCGAGTCT
TCGGACTCTTTGGTAACTGAGAGGCCCTAAGCATCAATCAAACGAAGTGAAGATAAACGCT
GAACCTGTTTCACAAGAACCAACAACCTCCAAGGGTTACTTCTCTCCCAAGCATCCCCCT
GTTTCGCCTCCTTGGCCTTACCAATACCCTCCAAACCCTAGTTTCTACCACATGCCCGTC
TACTGGGGCTGCGCGATACCGGTTTGGTCTACCTCGACACTTCTACATGTCTTGGGAAA
AGGACAAGAGACGAAACTTCTCATGAAACTGTTAAAGAGAGTAAAAATGCTTTTGAGAGA
ACAAGCTTGCTTTTGAATCTCAGAGCATCAAAAATGAAACAAGTATGGCTACAAATAAC
CATGTGTGGTATCCAGTACCGATGACCCGCGAGAAGACACAAGAATTACGCTTTTTTCAGT
AATGGAGCTGAAACAAAGAGCAGCAACAACAGATTCTGTCCTGAAACGTATCTTAACCTG
CAAGCAAACCTGCAGCCATGGCAAGATCTATGAACCTCAGAGAGAGCATATAA

>G1895 Amino Acid Sequence (domain in AA coordinates: 55-110)
MNNQSVTDNTSLKLSSNLNNEKSKTESSENSDDQHSEIITITSEEEKTTTELKPKDILPCPR
CNSADTKFCYNNYNVNQPRHFCRKCQRYWTAGGSMRIVPVGSGRRKNKGWVSSDQYLHI
TSEDTDNYSSTKILSFESSDSLVTTERPKHQSEVKINAEPVSQEPNPFQGLLPQASP
VSPFPWPYQYPNPSFYHMPVYWGCAIPVWSTLDTSTCLGKRTRDETSHETVKESKNAFER
TSLLESQSIKNETSMATNNHVWYVPVMTREKTQEFSSFGAETKSSNNRFPVETYLNL
QANPAAMARSMNFRSI*

>G1900 (1..897)

ATGCTGGAAACTAAAGATCCTGCGATAAAGCTCTTTGGTATGAAAATTCCTTTCCCGACG
GTTTTAGAGGTTGCTGATGAAGAAGAAAAGAACCAAAACAAGACATTAACTGATCAA
TCGGAGAAAGACAAACCCCTAAAGAAACCAACCAAGATTCTTCCATGTCCAAGATGCAAC
AGCATGGAGACTAAGTTCTGTTACTACAACAACCTACAACGTAAACCAACCTCGCCATTTT
TGTAAGCTTGTGATGAGATATTGGACCTCAGGTGGGACCATGAGAAGTGTTCGAATCGGA
GCAGGACGGCGCAAGAACAAGAACAACCTACCAACTTCACATTACCACCATGTGACTATC
TCCGAAACAAATGGTCCGGTCTTAGTTTCAGCCTCGGAGATGATCAAAAGGTCTCGAGT
AATAGGTTTGGTAATCAAAAGCTAGTTGCTAGGATAGAGAACAATGACGAGCGCTTAAT
AACAACACTTCGAACGGTTTGAATTGTTTCCGGGAGTTTCGTGGCCGTACACGTGGAAT
CCTGCGTTTACCCGTTTACCCTTATTGGAGCATGCCAGTGTGCTTCTCCGGTAAAGT
TCAAGTCTACTTCTACTCTTGGTAAGCATTCGAGAGACGAAGACGAGACGGTGAAGCAA
AAACAGAGGAATGGATCTGTATTGGTTCCAAAGACTTTGAGAATTGATGATCCTAATGAA

>G2007 Amino Acid Sequence (domain in AA coordinates: TBD)
MGRPCCDKLMVKKGPTAEEDKKLNFILTNHGHCWRALPKLAGLRRCGKSCRLRWITNY
LRPDLKRGLLSDAEAEQLVIDLHALLGNRWSKIAARLPGRITDNEIKNHWNTHIKKLLKME
IDPSTHQLPNKVFDTNLVDKSETSSKADNVNDKNIVEIDGTTTNTIDDSIITHQNSSND
DYELLGDIHNYGDLFNILWITNDEPLVDASNSHNKIVGIGTAAVAASDKNNTAAEEDF
PERSFEKQNGESWMFLDYCQEFVGVEDFGFECYHGFQSSMKTGHKD*

279

CGGTCTCGTGTGGCTCAAACACTCCGTCGAGTAGTGATGATGTTGAGGCGGATGCATCA
GAAAGGCAAGAGGATGGCACCAATGGTGAGGTGAAAGAAACGAATGAAGACACTAATAAA
CCTCAAACCTTCAGAGTCCAATGCACGCCGAGTAGAATCAGCTCCAATATAACCGATCCA
TGGAAGTCTGTGTCTGACGAGGGTCAATTGCCCTTCCAAGCTCTCTCTCCAGAGAGGTA
TTGCCGCAAAGTTTTACATATCGAGAAGAACACAGAGAGGAAGAACAACAACAAGAA
CAAAGATATCCAATGGCACTTGATCTTAACTTCACAGCTCAGTTAACACCAGTTGATGAT
CAAGAGGAGAAGAGAAACACAGGATTTCTTGAATCGGATTAGATGCTTCAAAGCTAATG
AGTAGAGGAAGAACAGGTTTTAAACCATACAAAAGATGTTCCATGGAAGCCAAAGAAAGT
AGAATCCTCAACAACAATCCTATCATTCATGTGGAACAGAAAGATCCCAAACGGATGCGG
TTGGAAACTCAAGCTTCCCATGAGACTCTATTTTCATCTGATCTGTTGTTTGTACTCTG
TTTTTAAGTTTTCAAGCACTGCTACATTTTCTTTTCTTTGAGGCCTTTGTATTTGT
TTCCTTGTCATAGTCTTCTGTAAACATTTGACTCTGTATTATTCAACAATCATAACT
GTTTAATCTTTTTTTTCCA

>G214 Amino Acid Sequence (domain in AA coordinates: 22-71)
METNSSGDELVIKTRKPYTITKQRRWTEEEHNRFI EALRLYGRAWQKIEEHVATKTAVQ
IRSHAQKFPFSKVEKEAEAKGVAMQALDIAIPPPRKRKPNPNPYPRKTGSGTILMSKTGV
NDGKESLGSEKVSHPMANEDRQSKPEEKLQEDNCSDCFTHQYLSAASSMNKSCIETS
NASTFREFLPSREEGSQNNRVRKESNSDLNAKSLENGNEQGQPYPMHIPVLVPLGSSIT
SSLSHPPSEPDSPHTVAGDYQSFNHHMSTLLQTPALYTAATFASSFWPPDSSGGSPVP
GNSPPNLAAMAAATVAASAWWAANGLLPLCAPLSSGGFTSHPPSTFGPSCDVEYTKAST
LQHGSVQSREQEHSEASKARSSLDSEVENKSKPVCHQPSATPESDAKSGDGAGDRKQV
DRSSCGSNTPSSDDVEADASERQEDGTNGEVKETNEDTNKPQTSSESNNRRSRISSNITD
PWKSVSDEGRIFALFQALFSREVLPSQFTYREEHREEEQQOEQRYPMALDLNFTAQLTPVD
DQEEKRNTGFLGIGLDASKLMSRGRGTGFKPKRCSMEAKESRILNNNPIIHVEQKDKPRM
RLETQAST*

>G2155 (63..740)

CTCATATATACCAACCAAACCTCTCTCTGCATCTTTATTAAACACAAAATTCAAAAGATT
AAATGTTGTGGAAGCTCCCTACACAGCGACACTTGCACCTCTCTCCCTCTCCCTCCA
TGGAACCGTCGGGCGTCCACGTGGCAGACCTCGAGGTTCCAAAAACAAACCTAAAGCTC
CAATCTTTGTCAACATTGACCCTCCTATGAGTCCTTACATCCTCGAAGTGCCATCCGGAA
ACGATGTCTGTTGAAGCCCTAAACCGTTTCTGCCGCGGTAAAGCCATCGGCTTTTGCCTCC
TCAGTGGCTCAGGCTCCGTTGCTGATGTCACCTTGCCTCAGCCTTCTCCGGCAGCTCCTG
GCTCAACCATTACTTTCCACGGAAGTTCGATCTTCTCTCTCTCTCCGCCACTTTCTCTCC
CTCCTCTACCTCCTACCTCCTTGTCCCTCCCGTCTCCAATTTCTTACCGTCTCTCTCG
CCGGACCTCAGGGGAAAGTCATCGGTGGATTGCTCGCTGGTCTCTCGTTGCCGCGGAA
CTGTTTACTTCGTCGCCACTAGTTTCAAGAACCCTTCTATCACCGGTTACCTGCTACGG
AGGAAGAGCAAAGAACTCGGCGGAAGGGGAAGAGGAGGGACAATCGCCGCCGGTCTCTG
GAGGTGGTGGGAGATCGATGTACGTGGGTGGCTCTGATGTCAATTGGGATCCCAACGCCA
AAGCTCCATCGCCGTACTGACCACAAATCCATCTCGTTCAAACCTAGGGTTTCTTCTTCTT
TAGATCATCAAGAATCAACAAAAGATTGCATTTTATAGATTCTTTGTAATATCATAATTG
ACTCACTCTTAACTCTCTATCACTTCTTCTTTAGCTTTTTCTGCAGTGTCAAACCTTCA
CATATTTGTAGTTTGATTGACTATCCCAAGTTTGTATTTTATCATACAAATTTTGC
CTGTCTCTAATGGTTGTTTTTCTGTTGTATAATCTTATGCATTGTTTATTGGAGCTCCA
GAGATTGAATGTATAATATAATGGTTAAT

>G2155 Amino Acid Sequence (domain in AA coordinates:18-38)
MLSKLPQRHLHLSPSSPSMETVGRPRGRPRGSKNPKAPIFVTIDPPMSPYILEVPSGN
DVVEALNRFRCRGAIGFCVLSGSGSVADVTLRQPSAAPGSTITFHGKFDLLSVSATFLP
PLPPTSLSPVSNFFTVSLAGPQGVIGGFVAGPLVAAGTVYFVATSFKNPSYHRLPATE
EEQRNSAEGEESQSPVSGGGGSMYVGGSDVIWDPNAKAPSPY*

>G234 (106..1035)

CACAACATCATACCCACCAACATATATAATCTTGATCATAGAGAGATAAACAGAGGCCGC
TATCAAGAAACAAGACTAAGAACAAAGACTTCACTAGGAGTACAAGTATGGGAAGAGCACCG
TGTTGTGACAAAGCAAACGTGAAGAAAGGGCCTTGGTCTCCTGAGGAAGATGCAAAACTC
AAATCTTACATTGAAAATAGTGGCACCGGAGGCAATTGGATCGCTTTGCCCTCAAAGATT
GGTTTTAAGAGATGTGGAAAGAGTTGCAGGCTGAGGTGGCTTAACTATCTTAGACCAAAC
ATCAAACATGGTGGCTTCTCTGAGGAAGAAGAAAACATATTGTAGCCTTTACCTTACA
ATTGGTAGCAGGTGGTCTATAATCGCTGCTCAATTGCCGGGACGAACAGACAACGATATA

AAAACTATTGGAACACGAGGCTCAAGAAGAACTCATTAACAAACAACGCAAGGAGCTT
CAAGAAGCTTGTATGGAGCAGCAAGAGATGATGGTGATGATGAAGAGACAACACCAACAA
CAACAAATCCAAACTTCTTTTATGATGAGACAAGACCAAAACAATGTTTACATGGCCACTA
CATCATCATAATGTTCAAGTTCAGCTCTTTTCAGAATCAAACCAACTCGTTTTGCGACC
AAGAAGATGTTAAGCCAGTGCATCAAGAACATGGTCAAGATCGAAGATCAAGAACTGG
AGAAAACAAACCTCATCATCAAGATTCAATGACAACGCTTTTGATCATCTCTCTTTC
TCTCAACTCTTGTAGATCCTAATCATAACCACCTTAGGATCAGGAGAGGGTTTCTCCATG
AACTCTATCTTGAGCGCCAACACAAACTCTCCATTGCTTAACACAAGTAATGATAATCAG
TGGTTCGGGAATTTCAGGCCGAAACCGTAAACTTGTCTCAGGAGCCTCCACAAGTACT
TCGGCAGATCAAAGCACTATAAGTTGGGAAGACATAAGCTCTCTGTTTATTCTGATTCA
AAGCAATTTTTTAATTATAATAATATATTATTCTTAAGATGAAACGTACATCATTATTA
TTAATTGGGGTACGTAACGTATATATGGAATAACGATCTAGTTTGTAAATTTAAAA
>G234 Amino Acid Sequence (domain in AA coordinates: 14-115)
MGRAPCCDKANVKKGPWSPEEDAKLKSIIENSGTGGNWIALPQKIGLKRCGKSCRLRWLN
YLRPNIKHGGFSEEEENIICSLYLTIGSRWSIIAQLPGRDNDIKNYWNTRLKKLINK
QRKELQEACMEQQEMVMKQHQQQQIQTSFMMRQDQTMFTWPLHHHNVQVPALFRIKP
TRFATKKMLSQCSSRTWSRSKIKNWRKQTSSSSRFNDNAFDHLSFSQLLDPNHNHLSGSG
EGFSMNSILSANTNSPLNLSNDNQWFGNFQAETVNLFSGASTSTSADQSTISWEDISSL
VYSDSKQFF*

>G361 (54..647)

TCTGTCTCTCTCTCTCTCTTTGTAAATATACATATATAGATAAGCTCACATATATGGCGA
CTGAAACATCTTCTTTGAAGCTCTTCGGTATAAACCTACTTGAAACGACGTCGGTTCAAA
ACCAGTCATCGGAACCAAGACCCGGATCCGGATCAGGATCCGAGTCACGTAAGTACGAGT
GTCAATACTGTTGTAGAGAGTTTGCTAACTCTCAAGCTCTTGGTGGTCACCAAAACGCTC
ACAAGAAAGAGCGTCAGCTTCTTAAACGTGCACAGATGTTAGCTACTCGTGGTTTGCCAC
GTCATCATAATTTTACCCTCATACCAATCCGCTTCTCTCCGCCCTTCGCGCCGCTGCCTC
ACCTCCTCTCTCAGCCGCATCCTCCGCCGCATATGATGCTCTCTCTCTCTCTCTCGAGTT
CTAAGTGGCTTTACGGTGAACACATGTCGTCACAAAACGCCGTTGGGTACTTTCATGGTG
GAAGGGGACTTTACGGAGGTGGCATGGAGTCTATGGCCGGAGAAGTAAAGACTCATGGTG
GTTCTTTGCCGGAGATGAGGAGTTTCGCCGGAGATAGTGATCGGAGTAGCGGAATTAAGT
TAGAGAATGGTATTGGGCTGGACCTCCATTTAAGCCTTGGGCCATGAATGATTATAATTT
TGGCCCAGTAAAGATCTGTAAAATACTACTAGGATTTTCATTTTATAGAGTATGTTTTTT
TCCTTAATTTCCGTTGAAATTGGTGAATATTTTATCTCTTACTTACCAATCTCATATT
TCTATGTATGCGTTTGTCTTTCATTTTTTTTTTATATAATTCTTCTGTAAAAAATGCA
ATGTGAGTTTTCTCCCTATCATTCTGTCAAGCTTTGGTTCAATTATTTAGTAATCGAAT
AATATAGGAATAGTGTGAAAG

>G361 Amino Acid Sequence (domain in AA coordinates: 43-63)
MATETSSSLKLFGINLLETTSVQNQSSEPRPGSGSGSESRYECQYCCREFANSQALGGHQ
NAHKKERQLLKRAQMLATRGLPRHNFHPHTNPLLSAFAPLPHLLSQPHPPHMLSPSS
SSSKWLYGEHMSSQNAVGYFHGGRGLYGGGMESMAGEVKTHGGSLEPMRRFAGDSDRSSG
IKLENGIGLDLHLSLGP*

>G562 (137..1285)

ATTTGAATTTCTGGGTTTCTCTCTGTTTAAAGCTTCTTCTTCTTCTCATCTTCTGCTTACGTT
TCTTCTTCAAGGAGCTTTCGGATTCTTGTAGAAAGAGTCATTGTTCTTCTTGGAGTGGGAAA
CCTTGAAACCATTCCTATGGGAAATAGCAGCGAGGAACCAAAGCCTCCTACCAATCAGA
TAAACCATCTTCACCCCGGTGGATCAACAAATGTTTATGTTCTACCTGATTGGGCAGC
TATGCAAGCATATTATGGTCCAAGAGTAGCAATGCCTCCTTATTACAATTACAGCTATGGC
TGCATCTGGTCATCTCTCTCTCTTACATGTGGAATCCTCAGCATATGATGTACCATC
TGGAGCACCCCTATGCTGCTGTTTATCTCATGAGGAGGAGTTTACGCTCATCCCGGTAT
TCCCATGGGATCACTGCCTCAAGGTCAAAGGATCCACCTTTAACAACCTCCGGGGACGCT
TTTGAGCATCGACACTCTACTAAATCTACAGGGAACACAGACAATGGATTGATGAAGAA
GCTGAAAGAGTTTGTATGGGCTTGTATGTCTCTAGGAAATGGGAATCCTGAAAATGGTGC
AGATGAACATAAACGATCAGGAACAGCTCAGAAACTGATGGTTCTACTGATGGAAGTGA
TGGGAATACAACCTGGGCGAGATGAACGAACTTAAAGAAGTCGAGAGGGAACCTCAAC
AAAAGATGGGAACAATGGTTTCAAGCTAGCTCATTTTATTCTGTTTCTCCGTCAAGTGG
TGATACCGGCGTAAACTCATTCAGGATCTGGAGCTATACTCTCTCTGTTGTAAGTGC
AAATTCACACCCCTCATGTCACAATCTTTAGCCATGGTTCTCTCTGAAACTTGCCTTCA

GAACGAGAGAGAACTGAAACGGGAGCGAAGGAAACAGTCTAATAGAGAATCTGCTAGAAG
GTCAAGATTAAGGAAACAGGCCGAGACAGAAGAACTTGCTAGGAAAGTGAAGCCTTGAC
AGCCGAAAACATGGCATTAAAGATCTGAACTAAACCAACTTAATGAGAAATCTGATAAACT
AAGAGGAGCAAATGCAACCTTGTGGACAACTGAAATGCTCGGAACCCGAAAAGAGAGT
CCCCGAAATATGTTGTCTAGAGTTAAGAACTCAGGAGCTGGAGATAAGAAACAAGAACCA
AGGAGACAATGATTCTAATCTACAAGCAAATTCATCAACTGCTCGATACGAAGCCTCG
AGCTAAAGCAGTAGCTGCAGGCTGAATCGATGGTAATTCATGTCGATTCTACTTAATTT
GTCGACATAAAAGAAAATAAGTGCTACTAATTTTCAGAAAACCTTGATAGATAGATAG
TATAGTAGAGAGAGAGAGAGAGAGAGAGGTGTGATGATTATTGATCTATAAATTTTCGGA
GAGAGAGAGGAGAGAAAGAGAACTTTTCTCCAGATGAAATTTGGTGTATGGTTTGT
ACTGTTAATATAGAGAGGCTTTTCTTTTATAAAATGGCTTCCTTTGTTGCA

>G562 Amino Acid Sequence (domain in AA coordinates: 253-315)
MGNSSSEPKPPTKSDKPSPPVDQTNVHVYPDWAAMQAYYGPVAMPYPYNSAMAASGHP
PPPYMWNPHMMSPSGAPYAAVYPHGGGVYAHGPIPMGSLPQGGKDPPLTTPGTLLSIDT
PTKSTGNTDNLMLKKLEFDGLAMSLGNGNPENGADHKRSRNSSETDGS TDGSDGNTTG
ADEPKLKRSTREGTPTKDGKQLVQASSFHSVSPSSGDTGVKLIQSGGAILSPGVANSNPF
MSQSLAMVPPETWLNRELRKRERKQSNRESARRSRLRKAETEELARKVEALTAENMA
LRSELNQLNEKSKLRGANATLLDLKLCSEPEKRVANMLSRVKNAGDKNKNQGDNDNS
NSTSKFHQLLDTKPRAKAVAAG*

>G591 (88..1020)

GTAAATCTCTCTTTGAAGGTTTCCTAACTCGTTAATCGTAACTCACAGTGACTCGTTTCGAG
TCAAAGTCTCTGTCTTTAGCTCAAACCATGGCTAGTAACAACCCTCACGACAACCTTTCT
GACCAAACCTCTTCTGATGATTTCTTCGAGCAAATCCTCGGCCTTCCTAACTTCTCAGCC
TCTTCTGCGCGCGGTTTATCTGGAGTTGACGGAGGATTAGGTGGTGGAGCACCGCCTATG
ATGCTGCAGTTGGGTTCCGGAGAGAAAGGAAGTCACATGGGTGGCTTAGGAGGAAGTGA
CCAACTGGGTTTACAATCAGATGTTTCTTTGGGGTTAAGTCTTGATCAAGGGAAGGA
CCTGGGTTTCTTAGACCTGAAGGAGACATGGAAGTGGGAAAAGATTCTCAGATGATGTT
GTTGATAATCGATGTTCTTCTATGAAACCTGTTTCCACGGGCAGCCTATGCAACAGCCA
CCTCCATCGGCCCCACATCAGCCTACTTCAATCCGTCCTCCAGGGTTCGAGCTAGGCGTGGT
CAGGCTACTGATCCACATAGCATCGCTGAGCGGCTACGTAGAGAAAGAATAGCAGAACGG
ATCAGGGCGCTGCAGGAACCTGTACCTACTGTGAACAAGACCGATAGAGCTGCTATGATC
GATGAGATTGTCGATTATGTAAAGTTTCTCAGGCTCCAAGTCAAGGTTTGGAGCATGAAC
CGACTTGGTGGAGCCGGTGCGGTTGCTCCACTTGTACTGATATGCCTCTTTCATCATCA
GTTGAGGATGAAACGGGTGAGGGTGAAGGACTCCGCAACCAGCGTGGGAGAAATGGTCT
AACGATGGGACTGAACGTCAGTGGCTAACTGATGGAAGAGAACGTTGGAGCCGCGATG
CAGCTTCTTCAATCAAAGGCTCTTTGTATGATGCCAATCTCATTTGGCAATGGCAATTTAC
CATCTCAACCTCCGGATACATCTTCAGTGGTCAAGCCTGAGAAACATCTCCACAGTAG
GATTTCTGCAATAAAGAGTTTGTACAGCTAATCCAACGTCCAACATGGGTTTTTCTTCT
GCTCTAATGACTCTGGTTTTCTTCTCTCTCCTCTCACCAGCTTGAAAGGTAAAAAAGTGAA
AAAGGCTTTGTAGATGGAATCAATGTAGGATTTCAGTAGAGGGCAAAAAATGTCATAT
AGCTCAATTGATCAAGTCTTAAAAA

>G591 Amino Acid Sequence (domain in AA coordinates: 143-240)
MASNNPHDNLSDQTPSDDFFEQILGLPNFSASSAAGLSGVDGGLGGGAPPMMLQLGSGBE
GSHMGLGSGSPTGFHNQMFPLGLSLDQKGPGFLRPEGGHSGSKRFSDVDVNDNRCSMK
PVFHGQPMQPPPSAPHQPTSIRPRVRARRGQATDPHSIAERLRERRERIAERIRALQELVP
TVNKTDRAMIDEIVDYVKFLRLQVKVLSMNLGGAGAVPLVTDMLSSSVEDETGE
RTPQPAWEKWSNDGTERQVAKLMEENVGAAMQLLQSKALCMMPISLAMAIYHSQPPDTSS
VVKPENPPQ*

>G8 (247..1596)

AAAAAAAATATCCGTCTCACTCTCTCGCCGCCGGTAACATTTCCCGCGACAAACTTC
TCTACTCTCACCATTCTCCATCGTAATCTCTAAATCTTCTCATTCTCTTCTCTCC
CGATCATCTCGAGCTCTCGTGAGAGATTATGTGATTATGTAATCGTTGTTGCTGTAGAA
GACGATCTCTAAACACTGATCTCTTCTCATCATCACCCTCGCTAGATTGTAATTTTCAGAG
CTTGAGATGTTGGATCTTAACCTCAACGCTGATTCTCCCGAGTCGACTCAGTACGGTGGT
GACTCATACTTAGATCGGCAGACATCAGACAACTCCGCCGGGAATCGAGTGAAGAGTCC
GGTACATCGACGTCGTAGTTATCAATGCCGATGGAGACGAAGACTCTTGCTCTACTCGA
GCTTTCACTCTCAGTTTTCGATATTTTAAAGTCGGAAGTAGTAGCGCGGAGACGAAAGC

MLDLNLNADSPSTQYGGDSYLDLRQTSNDSAGNRVRESGTSTSSVINADGDESDCSTRAF
TLSFDILKVGSSSGGDESPAASAVTKEFFPVSGDCGHLRDVEGSSSRNWIDLSFDRIG
DGETKLVTVPVTPAPVPAQVKKSRRGPRSSSQYRGVTFYRRTGRWESHIWDCGKQVYLG
GFDTAHAARAYDRAAIKFRGVDADINFTLGDYEEDMKQVQNLKEEFVHILRRQSTGFS
RGSSKRGVTLHKCGRWEARMGQFLGKKAYDKAAINTNGREAVTNFEMSSYQNEINSESN
NSEIDLNLGISLSTGNAPKONGRLFHFPSTYETQRGVSLRIDNEYMGKPVNTPLPYGSS
DHRLYWNGACSPSYNNPAEGRATEKRSEAEGMMSNWGWQRPGQTSAVRPQPPGPQPPPLFS
VAAASSGFSHFRPOPPNDNATRGYFPHP*

GATTTGTGCATTTTGTGTCTAGCCAAAAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAAACGAAGAAAAAGAACAAAAACATTGTGGGTCTCCGGTGATTAGGATCAAATTAGGGCACCAGCCTTATCGGAGGAAGAAGCCATGGGTAGAAAAAAGTCGAGATCAAGCGAATCGAGAACAAAGTAGTCGACAAGTCACTTTCTCCAAACGACGCAATGTCTCATCGAGAAAGCTCGACAACCTTCAATTCTCTGTGAATCTTCCATCGCTGTTCTCGTCGTCTCCGGCTCCGGAAAACTCTACAAGTCTGCCTCCGGTGACAACATGTCAAAGATCATGATCGTTACGAAATACATCATGCTGATGAACCTTGAAGCCTTAGATCTTGCAGAAAAAATTCGGAATTATCTGCCACTCAAGAGTTACTTAGAAATAGTCCAAAGCAAGCTTGAAGAATCAAATGTCTGATAATGTCAAGTGTGGTACTTTAATTTCTCTGAGGAAACAGCTCGAGACTGCTCTGTCCGTAAGTCTAGGAGCTAGGAAGACAGAACTAATGATGGGGGAAGTGAAGTCCCTTCAAAAAACGGAGAACTTGCTGAGAGAAGAGAACAGAGCTTTGGCTAGCCAGTGGGGAAGAAGACGTTTCTGGTTATAGAAGGTGACAGAGGAATGTCATGGGAAAATGGCTCCGGCAACAAGTACGGGAGACTCTCCGCTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCACCTTAAACTTACAGCCTGATTGAGAGTTTTCACAAATTTGTAAATTATAAAAAAGCTTCATAATAATCTCAACCTTTTATCTTCCCTCGCGCCAAATGTGGAAATTAAGGTTAAAAATAAAATAAACAGAGAGCTCATGCGAAAGAATTGTAAACTAAGATAAAGCTATAGTAGATCTTTATGTGTACCTTCGTAGCATATAAGATTTATTCGTGTGTTTGTCTTCCCTCNAAAAAAA
AAAAA

MGRRKVEIKRIENKSSRRQVTFSKRRNGLIEKARQLSILCESSIAVLVVSGSGKLYKSASG
DNMSKIIDRYEIIHHADEALDLAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL
EEQLETALSVTRARKTELMGGEVKS LQKTENLLREENQTLASQVGKKTFLVIEGDRGMSW
ENGSGNKVRETLPLLK*

CAAAAAAATCTCTCCCATTTAAAAGACTGCCCAAAGAAATATTTTATACAAAATGAAAGA
GAGAAACACGACACGAATTTTGTATAATTAAGATTACACAAAAAAAGTGTTAGAAAGAG

AAATATCTTCTTCTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTATCCTTTTGTTC
TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAGAACCATCGAAGTTAAATCATC
CACCGGAGTTTACAGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
TAGCGGTGGCGTTGGATTTAGTCCTGGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
TCCTGATGAGTTCAAGTCTTTCTCTCAGCTTTTAGCTGGAGCTATGGCTTCTCGGCGGC
AGCTGCTGTTGCCGCGCTGCTGTGGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
TGTCGGTGATGGCGGTGAAGCGGTGGTGATGTTGACCCGAGGTTTAAAGCAGAGTAGACC
AACGGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
GGCTACTCTTTTGGATTTCTCCGAGCTTCTTTGGTCTTTTTCACCTCTTCAGGGAACATT
TGGTATGACACATCAACAAGCTTTAGCACAAAGTCACTGCACAAGCAGTTCAAGGCAATAA
TGTTTCATATGCAGCAATCAACAATCTGAATATCCTTCTTCTACACAACAACAACA
ACAACAACAAGCTTTCATTGACTGAGATTCCATCATTTTTCTTCTGCACCTAGGTCTCA
GATTCGAGCCTCGGTTCAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAAATATCTGT
CTTTGAGCATCGGTTCACAGCCTCAAAATGCTGACAAACCAGCTGATGATGGATACAACCTG
GCGGAAATATGGGCAGAAGCAAGTGAAGGGGAGCGATTTTCTCGGAGTTATTACAAATG
TACGCATCCAGCTTGTCTGTCAAGAAGAAAGTGGAGAGGTCACTCGATGGACAAGTAAC
GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTTCAAAGCGCGGTAACAATAA
CGGGAGTTGTAAAGTTCTGATATTGCAAATCAGTTTCAAACAAGTAATAGCAGTCTCAA
CAAGAGTAAGAGGGACCAGGAAACAAGCCAAGTTACAACAACAGAGCAGATGTCTGAAGC
AAGTGATAGCGAGGAGGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
GCCTGATCCCAAGCGAAGAAATACAGAAGTTCCGGTTTCAGAACAGTTGCTTCATCGCA
TAGAACTGTGACAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
TGATGGATATAGTTGGCGCAAGTATGGTCAGAAAGTAGTCAAAGGAAATCCTTATCCGAG
GAGCTACTATAAGTGATGTAACAACAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAAACATAACCATGATGTTCCAGC
TGCTAGAACCAGCAGCCATCAGTTAAGACCAACAATCAACACAACACCTCAACGGTTAA
CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGCTTAAAGAAGAGCAAATCACTTGACA
GAGAAGAAGAAATACGACGGCGCTTGAGCTTTTGTGAGTTTAAATGAATCTTCTTTTGGTT
AATGAACCTGTTTTTGTGCTCAAAACACCACAGGTTTCTCTGGACAGAATCTCTGATA
TTACAGTTTCAAAGGTATGTTCTTTTATTTCATGTTGGAATCTTCTGTGTAATCTTAAG
AAGCTTTAGGAGGTAAATGTAACAAAACAGATTCAAAGTTATGCCCTTATGTGAATCTTT
TGACATGGGATAAACAATTTACAGGTATCCTTTTGTCTTGTGTAAAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKEEKEPSKLKSSSTGVSRTISLPPRPFGEMFFSGGVGFSPGPMTLVSNLFSDFDEFK
SFSQLLAGAMASPAAAVAAAVVATAHHQTPVSSVGDGGSGGDVDPRFKQSRPTGLMI
TQPPGMFTVPPGLSPATLLDSPSFFGLFSPLOQTFGMTHQQLAQVTAQAVQGNVHMQQ
SQQSEYPSSTQQQQQQQOQASLTEIPSFSSAPRSQIRASVQETSQOQRETSEISVFEHRS
QPQNADKPADDGYNWRKYQKQVKGSDFPRSYKCTHPACPVKKKVERSLDGQVTEIIYK
GQHNHELPQKRGNNNGSCKSSDIANQFQTSNSSLNKSQRDQETSQVTTTQMSSEASDSEE
VGNAETSVGERHEDEPDPRRNTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
RKYGQKVVKGPNPYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
HQLRPNQHNSTVNFNQQPVARLRLKEEQIT*

>G971 (131..1171)

TTTTTTTCTTCCCTCTTTTAGAATCTCTCTCTCTCTCGTTTTTGACACTTATCCTCTC
TCTTTTTTCTCTCTCCCTCTCTCTCTGGCCGGAAGAAACAACGTCGTTTATAGCTAA
AGATTCGATCATGTTGGATCTTAACCTAAAGATCTTTTCTTCTTATAACGAAGATCAAGA
TCGGAAAGTACCATTATGATCTCAACCACCGGTGAAGAAGAATCTAACTCATCTTCCTC
CTCCACAACAGACTCTGCAGCGAGAGATGCTTTTCATCGCTTTTGGAAATCTCAAACGCGA
CGATGACCTTGTTCTCTCTCTCTCTCTCTCTCATAAGAAACAGGAGATCTCTTCC
GGTGGTGGCTGATGCTCGTGGGAATATAGAATCTCCGTGAAGACAGTCACTGGTTGAA
TCTTTCTCTTTTACAAAGAAATACACAGAAATGGTGAAGAAGAGCAGAAGAGGACCAAG
GTCTCGTAGCTCCCAATATCGTGGCGTCACTTTTACCCTCGCACCGGTCTGTTGGGAATC
TCATATTTGGGATTGTGGAAAGCAAGTTTATTTGGGCGGGTTGATACTGCTTACGCAGC
AGCAAGGGCTTACGACCGAGCTGCTATCAAATTCGGTGGTCTCGATGCAGACATCAATTT
CGTCTGGATGATTATAGGCATGACATCGATAAGATGAAGAATTTAAATAAGGTGGAGTT
CGTGCAAAACACTTAGGCGAGAGAGTGCGAGTTTTCGGAAGAGGAAGTTCCAAATACAAAGG

CTTGGCTCTTCAAAAATGCACCCAATTCAAAACTCATGATCAGATTCATCTCTTCCAAAA
CAGGGGATGGGATGCAGCAGCAATAAAATACAATGAGTTGGGAAAGGGAGAAGGAGCCAT
GAAGTTTGGTGCCCATATCAAAGGAAATGGTCACAATGATCTTGAACCTAGTCTCGGAAT
TTCATCATCATCGGAAAGTATAAAGTTGACAACAGGCGATTACTATAAGGGTATCAATCG
GTCCACGATGGGTTTATACGGTAAGCAATCATCGATATTTTACCCATGGCAACCATGAA
ACCTCTGAAGACAGTTGACGATCATCAGGATTCCTTTTATCAGCATGACAAGTTCCTC
TTCTCCATGTCCAATTGTTTGTATCCATAGGATCGTTCTACACTCTCTTAACCTAATATA
TATTTTTACTCTATCTGATTATTGTATACAAGGATAAAATTTGATTCTTTCCTTAATGAG
TGAGAAATATTGGAAGTGTTAAAAA

>G971 Amino Acid Sequence (conserved domain in aa coordinates: 120-186)

MLDLNLKIFSSYNEDQDRKVPMLMISTTGEESNSSSSSTTDSAADAFIAFGILKRDDDL
VPPPPPPPHKETGDLFPVADARNNIEFVSDSHWLNLSLQNRNTQKMKVKSRRGPRSR
SQYRGVTFYRRRTGRWESHIWDCGKQVYLGFDYAAAAAYDRAAIKFRGLDADINFVVD
DYRHDIDKMKNLNKVEFVQTLRRRESASFGRGSSKYKGLALQKCTQFKTHDQIHLFQNRGW
DAAAIKYNELGKGEGAMKFGAHIKNGHNDLELSLGISSSSSESILKTTGDYYKGINRSTM
GLYKQSSIFLPMATMKPLKTVASSGFPFISMTSSSSSSMSNCFDP*

>G975 (58..657)

ATTACTCATCATCAAGTTCTCTCTCTGACAAACATCACAGAGTAAGTAAGAATG
GTACAGACGAAGAAGTTCAGAGGTGTGAGGCAACGCCATTGGGGTCTTGGGTCGCTGAG
ATTCGTCTCTCTCTTGAACGGAGGATTGGCTAGGGACGTTTCGAGACCGCAGAGGAG
GCAGCAAGACATACGACGAGGCCCGCTTTAATGAGCGGCCGCAACGCCAAAACCAAC
TTTCCCCTCAACAACAACAACACCGGAGAACTTCCGAGGGCAAAACCGATATTTCAGCT
TCGTCCACAATGTCTCTCAACATCATCTTCATCGCTCTCTTCCATCCTCAGCGCCAAA
CTGAGGAAATGCTGCAAGTCTCTTCCCCTACCTGCCTCCGTCTTGACACAGCC
AGCTCCCATATCGGCGTCTGGCAGAAACGGGCGGTTCAAAGTCTGACTCCAGCTGGGTC
ATGACGGTGGAGCTAGGTCCCGAAGCTCCTCCCAAGAGACTACTAGTAAAGCTTCACAA
GACGCTATTCTTGCTCCGACCTGAAGTTGAAATTGGTGGCAGCAGAGAAGAAGTATTG
GATGAGGAAGAAAAGGTTGCTTTGCAAATGATAGAGGAGCTTCTCAATACAACTAAATC
TTATTGCTTATATATATGTACCTATTTTCATTGCTGATTTACAGCCAAAATAATCAATT
ATACCGTGATTTTATAGATGTTTTATATTAAAGGTTGTTAGATATA

>G975 Amino Acid Sequence (domain in AA coordinates: 4-71)

MVQTKKFRGVRQRHWGWSVAEIRHPLLRRIWLGTFTAEAAARAYDEAAVLMSGRNAKT
NFPLNNNTGETSEGKTDISASSTSSSSLSLSAKLRKCKKSPSPSLTCLRLDT
ASSHIGVWQKRAGSKDSSWMTVELGPASSSQETTSKASQDAILAPTTEVEIGGSREEV
LDEEEKVALQMIEELLNTN*

>G994 (180..917)

TGTATATATAGTTAGTTAGTTGAGATAAACTTGGTTACCACTTTTGTGTGGTCTTTCTTT
TTCTTTTCTCCATTTTCCATTTATCGACCCCTTGGGTGTAGCTAATTACTTTTCGCGATT
TTCAAATCCAATAAAGTTTAATTTGATGAAGCTTTTAAACCATATAATATAAATAA
TGGGTGGTTCGTAACCATGTTGTGTATGAGGTTGGATTAAAGAAAGGTTCCATGGACAGTGG
AAGAAGATGGGAAACTAGTTGATTTCTTAAGGGCACGTGGCAACTGCGGTGGTGGTGGAG
GAGGATGGTGTGGAGAGACGTGCCAAAACCTGGCGGGGCTAAGGAGGTGTGGCAAAAGTT
GCCGTCTCCGGTGGACTAATTATCTCCGGCCAGATCTCAAGAGAGGTCTTTTACTGAAG
AAGAAATCCAACCTAGTCATTGATCTTCATGCTCGCCTTGGCAATAGATGGTTCGAAGATTG
CAGTGGAGTTACCAGGAAGAACAGACAACGATATCAAAAATTATTGGAACACTCATATAA
AGAGGAAGCTTATAAGAAATGGGTATTGATCCAAACACACATCGTCGATTTGACCAACAAA
AAGTCAACGAGGAGGAAACGATATTGGTCAACGATCCAAAGCCTCTGTCTGAGACCGAGG
TATCTGTGCTTTGAAGAATGACACGTGAGCAGTGTATCAGGAAATCTAAACCAATTGG
CTGACGTGGACGGTGTATGATCAGCGTGGAGCTTTCTAATGGAAAATGACGAAGGAGGAG
GTGGCGACGCCCGGAGAGCTTACGATGCTATTGTCCGGTGACATTACGTCATCATGTT
CTTCTTCGTCTCTTTGTGGATGAAGTATGGAGAATTTCGGATACGAAGATTTAGAAGTTG
GATGTTTCGATGTTTATAGATTCAAGTATGTTTAATTAGGCCGTAGGTTGATTAATCATA
AGGTTTCATTGACTTCATTCTAGAAATGTGTAGTTGGACCAGTATAAGAATCAAAGTTAT
GAAACATTGTAATTTGATTTCCAAATTAATCTAATGAATAAATGTGCTTTGCAAAAAA
AAAAA

>G994 Amino Acid Sequence (domain in AA coordinates: 14-123)
MGGRKPCDEVGLRKGPWTVEDGKLVDFLRARGNCGGGGGWCWRDVPKLAGLRRCGKS

CRLRWNTNYLRPDLKRGLFTEEEIQLVIDLHARLGNRWSKI AVELPGRTDNDIKNYWNTHI
KRKLIRMGIDPNTHRRFDQQKVNEEETILVNDPKPLSETEVSVALKNDTS AVLSGNLNQL
ADV DGGDQ PWSFLMENDEGGGGDAAGELTMLLSGDITSSCSSSSSLWMKYGEFGYEDLEL
GCFDV*

>G2347 (81..626)

AGCCCATCCTTCAACATTGCTTCCTAACCAGAAATCCACCATCATCTTCCCACGAATACA
ACTTAAAGCTTTACAGAAAATGGAGGGTCAGAGAACAACGCCGGGGTTACTTGAAAG
ACAAGGCTACAGTCTCCAACCTTGTGGAAGAAGAAATGGAGAATGGCATGGATGGAGAAG
AGGAGGATGGAGGAGACGAAGACAAAAGGAAGAAGGTGATGGAAAGAGTTAGAGGTCCTA
GCACTGACCGTGTTCATCGCGACTGTGCCAGGTCGATAGGTGCACTGTTAATTTGACTG
AGGCCAAGCAGTATTACCCGACAGACACAGAGTATGTGAAGTACATGCAAAGGCATCTGCTG
CGACTGTTGCAGGGGTCAGGCAACGCTTTTGTCAACAATGCAGCAGGTTTCATGAGCTAC
CAGAGTTTGTGAAGCTAAAAGAAGCTGCAGGAGGCGCTTAGCTGGACACAATGAGAGGA
GGAGGAAGATCTCTGGTGACAGTTTTGGAGAAGGGTCAGGCCGGAGAGGGTTTAGCGGTC
AACTGATCCAGACTCAAGAAAGAAACAGGGTAGACAGGAACTTCCTATGACCAACTCAT
CATTCAAGCGACCACAGATCAGATAAACCTCCCGCTCTCTCTCTTCTGTCTATCTACATA
TGCTCTATCTACACTCTTATTAGACAAATAATGGCATCTAACAATGTCAAGAAAAGTTGG
TCATGGTATTAAATCCTACCGGATATATACTATAAACTCTAGTCCCTCTATGCTGT
CCTGTAATGAATATCTATCCGAAATGTATTGCGCATAGTCTTGCGTCTAATAATGTTTAT
TGATTTTGTA

>G2347 Amino Acid Sequence (domain in AA coordinates: 60-136)

MEGQRTQRRGYLKDKATVSNLVEEEMENGMDGEEEDGGDEDKRKKVMERVGRGPSTDRVPS
RLCQVDRCTVNLTEAKQYYRRHRVCEVHAKASAATVAGVRQRFCCQCSRFFHELPEFDEAK
RSCRRRLAGHNERRRKISGDSFGEESGRRGFSGQLIQTQERNRVRDKLPMTNSSFKRPQI
R*

>G2010 (1..525)

ATGGAGGGTAAGAGATCACAAGGACAAGGTTACATGAAAAAGAAGTCTTACCTTGTGGAA
GAAGATATGGAGACTGATACGGATGAAGAAGAGGAAGTAGGTAGGGATAGAGTTAGAGGG
TCTAGAGGTAGCATCAATCGTGGTGGCTCGTTGCGGCTTTGCCAAGTAGATAGATGCACA
GCTGATATGAAAGAGGCAAACTGTATCACCGGAGACACAAAGTGTGTGAAGTTCATGCA
AAGGCATCTTCTGTCTTTCTCTCAGGACTTAACCAACGCTTTTGTCAACAATGCAGTAGG
TTTCATGACCTCCAAGAGTTTGATGAAGCTAAGAGAAGTTGCAGGAGGCGCTTAGCTGGA
CACAATGAGCGAAGAAGGAAGAGCTCTGGTGAGAGTACTTATGGAGAAGGATCAGGTCGG
AGAGGAATCAATGGTCAGGTGGTGTATGCAGAATCAAGAAAGATCAAGGGTAGAGATGACA
CTTCCTATGCCAAACTCATATTCAAGCGACCACAGATTAGATAG

>G2010 Amino Acid Sequence (domain in AA coordinates: 53-127)

MEGKRSQGGQYMKKKSYLVEEDMETDTDEEEVGRDRVRGSRGSRINRGSLRLCQVDRCT
ADMKEAKLYHRRHKVCEVHAKASSVFLSGLNQRFCQCSRFDLQEFDEAKRSCRRRLAG
HNERRRKSSGESTYGEESGRRGNGQVVMQNQERSRVEMTLMPNPSSFKRPQIR*

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
20 February 2003 (20.02.2003)

PCT

(10) International Publication Number
WO 03/013227 A2

(51) International Patent Classification⁷: A01H

(US). YU, Guo-Liang [US/US]; 242 Gravett Drive, Berkeley, CA 94705-1531 (US). BROUN, Pierre, E. [FR/US]; 921 Sunnybrae Boulevard, San Mateo, CA 94402 (US).

(21) International Application Number: PCT/US02/25805

(22) International Filing Date: 9 August 2002 (09.08.2002)

(74) Agents: WARD, Michael, R. et al.; Morrison & Foerster LLP, 425 Market Street, San Francisco, CA 94105-2482 (US).

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/310,847	9 August 2001 (09.08.2001)	US
60/336,049	19 November 2001 (19.11.2001)	US
60/338,692	11 December 2001 (11.12.2001)	US
10/171,468	14 June 2002 (14.06.2002)	US

(81) Designated States (*national*): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(71) Applicant (*for all designated States except US*): MENDEL BIOTECHNOLOGY, INC. [US/US]; 21375 Cabot Boulevard, Hayward, CA 94545 (US).

(84) Designated States (*regional*): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): RATCLIFFE, Oliver [GB/US]; 814 East 21st Street, Oakland, CA 94606 (US). RIECHMANN, Jose, Luis [ES/US]; 518 S. El Molino Avenue, #308, Pasadena, CA 91101 (US). ADAM, Luc, J. [CA/US]; 25800 Industrial Boulevard, Apt. L403, Hayward, CA 94545 (US). DUBELL, Arnold, N. [US/US]; 14857 Wake Avenue, San Leandro, CA 94578 (US). HEARD, Jacqueline, E. [US/US]; 810 Guilford Avenue, San Mateo, CA 94402 (US). PILGRIM, Marsha, L. [US/US]; 1368 Patrick Henry Drive, Phoenixville, PA 19460 (US). JIANG, Cai-Zhong [US/US]; 34495 Heathrow Terrace, Fremont, CA 94555 (US). REUBER, T., Lynne [US/US]; 1115 S. Grant Street, San Mateo, CA 94402 (US). CREELMAN, Robert, A. [US/US]; 2801 Jennifer Drive, Castro Valley, CA 94546 (US). PINEDA, Omaira [CO/US]; 4060 9th Place, Vero Beach, FL 32960

Published:

— without international search report and to be republished upon receipt of that report

(48) Date of publication of this corrected version:

3 July 2003

(15) Information about Correction:

see PCT Gazette No. 27/2003 of 3 July 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

(57) Abstract: The invention relates to plant transcription factor polypeptides, polynucleotides that encode them, homologs from a variety of plant species, and methods of using the polynucleotides and polypeptides to produce transgenic plants having advantageous properties compared to a reference plant. Sequence information related to these polynucleotides and polypeptides can also be used in bioinformatic search methods and is also disclosed.

WO 03/013227 A2

YIELD-RELATED POLYNUCLEOTIDES AND POLYPEPTIDES IN PLANTS

This application claims the benefit of US Provisional Application No. 60/310,847, filed August 9, 2001, US Provisional Application No. 60/336,049, filed December 5, 2001, US Provisional Application No. 60/338,692, filed December 11, 2001, and US Non-provisional Application No. 10/171,468, filed June 14, 2002, the entire contents of which are hereby incorporated by reference.

FIELD OF THE INVENTION

This invention relates to the field of plant biology. More particularly, the present invention pertains to compositions and methods for phenotypically modifying a plant.

INTRODUCTION

A plant's traits, such as its biochemical, developmental, or phenotypic characteristics, may be controlled through a number of cellular processes. One important way to manipulate that control is through transcription factors - proteins that influence the expression of a particular gene or sets of genes. Transformed and transgenic plants that comprise cells having altered levels of at least one selected transcription factor, for example, possess advantageous or desirable traits. Strategies for manipulating traits by altering a plant cell's transcription factor content can therefore result in plants and crops with commercially valuable properties. Applicants have identified polynucleotides encoding transcription factors, developed numerous transgenic plants using these polynucleotides, and have analyzed the plants for a variety of important traits. In so doing, applicants have identified important polynucleotide and polypeptide sequences for producing commercially valuable plants and crops as well as the methods for making them and using them. Other aspects and embodiments of the invention are described below and can be derived from the teachings of this disclosure as a whole.

BACKGROUND OF THE INVENTION

Transcription factors (TFs) can modulate gene expression, either increasing or decreasing (inducing or repressing) the rate of transcription. This modulation results in differential levels of gene expression at various developmental stages, in different

tissues and cell types, and in response to different exogenous (e.g., environmental) and endogenous stimuli throughout the life cycle of the organism.

Because transcription factors are key controlling elements of biological pathways, altering the expression levels of one or more transcription factors can change entire biological pathways in an organism. For example, manipulation of the levels of selected transcription factors may result in increased expression of economically useful proteins or metabolic chemicals in plants or to improve other agriculturally relevant characteristics. Conversely, blocked or reduced expression of a transcription factor may reduce biosynthesis of unwanted compounds or remove an undesirable trait. Therefore, manipulating transcription factor levels in a plant offers tremendous potential in agricultural biotechnology for modifying a plant's traits.

The present invention provides novel transcription factors useful for modifying a plant's phenotype in desirable ways.

SUMMARY OF THE INVENTION

In a first aspect, the invention relates to a recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of: (a) a nucleotide sequence encoding a polypeptide comprising a polypeptide sequence selected from those of the Sequence Listing, SEQ ID NOs:2 to 2N, where $N = 2-561$, or those listed in Table 4, or a complementary nucleotide sequence thereof; (b) a nucleotide sequence encoding a polypeptide comprising a variant of a polypeptide of (a) having one or more, or between 1 and about 5, or between 1 and about 10, or between 1 and about 30, conservative amino acid substitutions; (c) a nucleotide sequence comprising a sequence selected from those of SEQ ID NOs:1 to $(2N - 1)$, where $N = 2-561$, or those included in Table 4, or a complementary nucleotide sequence thereof; (d) a nucleotide sequence comprising silent substitutions in a nucleotide sequence of (c); (e) a nucleotide sequence which hybridizes under stringent conditions over substantially the entire length of a nucleotide sequence of one or more of: (a), (b), (c), or (d); (f) a nucleotide sequence comprising at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides of a sequence of any of (a)-(e), or at least 10 or 15, or at least about 20, or at least about 30 consecutive nucleotides outside of a region encoding a conserved domain of any of (a)-(e); (g) a

nucleotide sequence comprising a subsequence or fragment of any of (a)-(f), which subsequence or fragment encodes a polypeptide having a biological activity that modifies a plant's characteristic, functions as a transcription factor, or alters the level of transcription of a gene or transgene in a cell; (h) a nucleotide sequence having at least 31% sequence identity to a nucleotide sequence of any of (a)-(g); (i) a nucleotide sequence having at least 60%, or at least 70 %, or at least 80 %, or at least 90 %, or at least 95 % sequence identity to a nucleotide sequence of any of (a)-(g) or a 10 or 15 nucleotide, or at least about 20, or at least about 30 nucleotide region of a sequence of (a)-(g) that is outside of a region encoding a conserved domain; (j) a nucleotide sequence that encodes a polypeptide having at least 31% sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; (k) a nucleotide sequence which encodes a polypeptide having at least 60%, or at least 70 %, or at least 80%, or at least 90 %, or at least 95 % sequence identity to a polypeptide listed in Table 4, or the Sequence Listing; and (l) a nucleotide sequence that encodes a conserved domain of a polypeptide having at least 85%, or at least 90%, or at least 95%, or at least 98% sequence identity to a conserved domain of a polypeptide listed in Table 4, or the Sequence Listing. The recombinant polynucleotide may further comprise a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence. The invention also relates to compositions comprising at least two of the above-described polynucleotides.

In a second aspect, the invention comprises an isolated or recombinant polypeptide comprising a subsequence of at least about 10, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids encoded by the recombinant or isolated polynucleotide described above, or comprising a subsequence of at least about 8, or at least about 12, or at least about 15, or at least about 20, or at least about 30 contiguous amino acids outside a conserved domain.

In a third aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fourth aspect, the invention comprises an isolated or recombinant polynucleotide that encodes a polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a fifth aspect, the invention comprises an isolated polypeptide that is a paralog of the isolated polypeptide described above. In one aspect, the invention is an paralog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

In a sixth aspect, the invention comprises an isolated polypeptide that is an ortholog of the isolated polypeptide described above. In one aspect, the invention is an ortholog which, when expressed in *Arabidopsis*, modifies a trait of the *Arabidopsis* plant.

The present invention also encompasses transcription factor variants. A preferred transcription factor variant is one having at least 40% amino acid sequence identity, a more preferred transcription factor variant is one having at least 50% amino acid sequence identity and a most preferred transcription factor variant is one having at least 65% amino acid sequence identity to the transcription factor amino acid sequence SEQ ID NOs:2 to 2N, where $N = 2-561$, and which contains at least one functional or structural characteristic of the transcription factor amino acid sequence. Sequences having lesser degrees of identity but comparable biological activity are considered to be equivalents.

In another aspect, the invention is a transgenic plant comprising one or more of the above-described isolated or recombinant polynucleotides. In yet another aspect, the invention is a plant with altered expression levels of a polynucleotide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a nucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, *Arabidopsis*, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry,

raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In yet another aspect the invention is a transgenic plant comprising one or more of the above described polynucleotides wherein the encoded polypeptide is expressed and regulates transcription of a gene.

In a further aspect the invention provides a method of using the polynucleotide composition to breed a progeny plant from a transgenic plant including crossing plants, producing seeds from transgenic plants, and methods of breeding using transgenic plants, the method comprising transforming a plant with the polynucleotide composition to create a transgenic plant, crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

In a further aspect, the invention provides a progeny plant derived from a parental plant wherein said progeny plant exhibits at least three fold greater messenger RNA levels than said parental plant, wherein the messenger RNA encodes a DNA-binding protein which is capable of binding to a DNA regulatory sequence and inducing expression of a plant trait gene, wherein the progeny plant is characterized by a change in the plant trait compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least ten fold greater messenger RNA levels compared to said parental plant. In yet a further aspect, the progeny plant exhibits at least fifty fold greater messenger RNA levels compared to said parental plant.

In a further aspect, the invention relates to a cloning or expression vector comprising the isolated or recombinant polynucleotide described above or cells comprising the cloning or expression vector.

In yet a further aspect, the invention relates to a composition produced by incubating a polynucleotide of the invention with a nuclease, a restriction enzyme, a polymerase; a polymerase and a primer; a cloning vector, or with a cell.

Furthermore, the invention relates to a method for producing a plant having a modified trait. The method comprises altering the expression of an isolated or recombinant polynucleotide of the invention or altering the expression or activity of a polypeptide of the invention in a plant to produce a modified plant, and selecting the modified plant for a modified trait. In one aspect, the plant is a monocot plant. In another aspect, the plant is a dicot plant. In another aspect the recombinant polynucleotide is from a dicot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a dicot plant. In yet another aspect the recombinant polynucleotide is from a monocot plant and the plant is a monocot plant. In yet another aspect the recombinant polynucleotide is from a dicot plant and the plant is a dicot plant.

In another aspect, the invention is a transgenic plant comprising an isolated or recombinant polynucleotide encoding a polypeptide wherein the polypeptide is selected from the group consisting of SEQ ID NOs: 2 - 2N, where $N = 2-561$. In yet another aspect, the invention is a plant with altered expression levels of a polypeptide described above or a plant with altered expression or activity levels of an above-described polypeptide. Further, the invention is a plant lacking a polynucleotide sequence encoding a polypeptide described above or substantially lacking a polypeptide described above. The plant may be any plant, including, but not limited to, Arabidopsis, mustard, soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits, vegetable brassicas, and mint or other labiates. In yet another aspect, the inventions is an isolated plant material of a plant, including, but not limited to, plant tissue, fruit, seed, plant cell, embryo, protoplast, pollen, and the like. In yet another aspect, the

invention is a transgenic plant tissue culture of regenerable cells, including, but not limited to, embryos, meristematic cells, microspores, protoplast, pollen, and the like.

In another aspect, the invention relates to a method of identifying a factor that is modulated by or interacts with a polypeptide encoded by a polynucleotide of the invention. The method comprises expressing a polypeptide encoded by the polynucleotide in a plant; and identifying at least one factor that is modulated by or interacts with the polypeptide. In one embodiment the method for identifying modulating or interacting factors is by detecting binding by the polypeptide to a promoter sequence, or by detecting interactions between an additional protein and the polypeptide in a yeast two hybrid system, or by detecting expression of a factor by hybridization to a microarray, subtractive hybridization, or differential display.

In yet another aspect, the invention is a method of identifying a molecule that modulates activity or expression of a polynucleotide or polypeptide of interest. The method comprises placing the molecule in contact with a plant comprising the polynucleotide or polypeptide encoded by the polynucleotide of the invention and monitoring one or more of the expression level of the polynucleotide in the plant, the expression level of the polypeptide in the plant, and modulation of an activity of the polypeptide in the plant.

In yet another aspect, the invention relates to an integrated system, computer or computer readable medium comprising one or more character strings corresponding to a polynucleotide of the invention, or to a polypeptide encoded by the polynucleotide. The integrated system, computer or computer readable medium may comprise a link between one or more sequence strings to a modified plant trait.

In yet another aspect, the invention is a method for identifying a sequence similar or homologous to one or more polynucleotides of the invention, or one or more polypeptides encoded by the polynucleotides. The method comprises providing a sequence database, and querying the sequence database with one or more target sequences corresponding to the one or more polynucleotides or to the one or more polypeptides to identify one or more sequence members of the database that display sequence similarity or homology to one or more of the one or more target sequences.

The method may further comprise of linking the one or more of the polynucleotides of the invention, or encoded polypeptides, to a modified plant phenotype.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING, TABLES, AND FIGURE

The Sequence Listing provides exemplary polynucleotide and polypeptide sequences of the invention. The traits associated with the use of the sequences are included in the Examples.

Diskette1 is a read-only memory computer-readable diskette and contains a copy of the Sequence Listing in ASCII text format. The Sequence Listing is named "SEQLIST514442002041" and is 929 kilobytes in size. The copy of the Sequence Listing on the diskette is hereby incorporated by reference in its entirety.

Table 4 shows the polynucleotides and polypeptides identified by SEQ ID NO; Mendel Gene ID No.; conserved domain of the polypeptide; and if the polynucleotide was tested in a transgenic assay. The first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Table 5 lists a summary of orthologous and homologous sequences identified using BLAST (tblastx program). The first column shows the polynucleotide sequence identifier (SEQ ID NO), the second column shows the corresponding cDNA identifier (Gene ID), the third column shows the orthologous or homologous polynucleotide GenBank Accession Number (Test Sequence ID), the fourth column shows the

calculated probability value that the sequence identity is due to chance (Smallest Sum Probability), the fifth column shows the plant species from which the test sequence was isolated (Test Sequence Species), and the sixth column shows the orthologous or homologous test sequence GenBank annotation (Test Sequence GenBank Annotation).

Figure 1 shows a phylogenic tree of related plant families adapted from Daly et al. (2001 *Plant Physiology* 127:1328-1333).

Detailed Description of Exemplary Embodiments

In an important aspect, the present invention relates to polynucleotides and polypeptides, e.g. for modifying phenotypes of plants. Throughout this disclosure, various information sources are referred to and/or are specifically incorporated. The information sources include scientific journal articles, patent documents, textbooks, and World Wide Web browser-inactive page addresses, for example. While the reference to these information sources clearly indicates that they can be used by one of skill in the art, applicants specifically incorporate each and every one of the information sources cited herein, in their entirety, whether or not a specific mention of "incorporation by reference" is noted. The contents and teachings of each and every one of the information sources can be relied on and used to make and use embodiments of the invention.

It must be noted that as used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a plant" includes a plurality of such plants, and a reference to "a stress" is a reference to one or more stresses and equivalents thereof known to those skilled in the art, and so forth.

The polynucleotide sequences of the invention encode polypeptides that are members of well-known transcription factor families, including plant transcription factor families, as disclosed in Table 4. Generally, the transcription factors encoded by the present sequences are involved in cell differentiation and proliferation and the regulation of growth. Accordingly, one skilled in the art would recognize that by expressing the present sequences in a plant, one may change the expression of

autologous genes or induce the expression of introduced genes. By affecting the expression of similar autologous sequences in a plant that have the biological activity of the present sequences, or by introducing the present sequences into a plant, one may alter a plant's phenotype to one with improved traits. The sequences of the invention may also be used to transform a plant and introduce desirable traits not found in the wild-type cultivar or strain. Plants may then be selected for those that produce the most desirable degree of over- or underexpression of target genes of interest and coincident trait improvement.

The sequences of the present invention may be from any species, particularly plant species, in a naturally occurring form or from any source whether natural, synthetic, semi-synthetic or recombinant. The sequences of the invention may also include fragments of the present amino acid sequences. In this context, a "fragment" refers to a fragment of a polypeptide sequence which is at least 5 to about 15 amino acids in length, most preferably at least 14 amino acids, and which retain some biological activity of a transcription factor. Where "amino acid sequence" is recited to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

As one of ordinary skill in the art recognizes, transcription factors can be identified by the presence of a region or domain of structural similarity or identity to a specific consensus sequence or the presence of a specific consensus DNA-binding site or DNA-binding site motif (see, for example, Riechmann et al., (2000) *Science* 290: 2105-2110). The plant transcription factors may belong to one of the following transcription factor families: the AP2 (APETALA2) domain transcription factor family (Riechmann and Meyerowitz (1998) *Biol. Chem.* 379:633-646); the MYB transcription factor family (Martin and Paz-Ares, (1997) *Trends Genet.* 13:67-73); the MADS domain transcription factor family (Riechmann and Meyerowitz (1997) *Biol. Chem.* 378:1079-1101); the WRKY protein family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the ankyrin-repeat protein family (Zhang et al. (1992) *Plant Cell* 4:1575-1588); the zinc finger protein (Z) family (Klug and Schwabe (1995) *FASEB J.* 9: 597-604); the homeobox (HB) protein family (Buerklin in *Guidebook to the Homeobox Genes*, Duboule (ed.) (1994) Oxford University Press);

the CAAT-element binding proteins (Forsburg and Guarente (1989) *Genes Dev.* 3:1166-1178); the squamosa promoter binding proteins (SPB) (Klein et al. (1996) *Mol. Gen. Genet.* 1996 250:7-16); the NAM protein family (Souer et al. (1996) *Cell* 85:159-170); the IAA/AUX proteins (Rouse et al. (1998) *Science* 279:1371-1373); the HLH/MYC protein family (Littlewood et al. (1994) *Prot. Profile* 1:639-709); the DNA-binding protein (DBP) family (Tucker et al. (1994) *EMBO J.* 13:2994-3002); the bZIP family of transcription factors (Foster et al. (1994) *FASEB J.* 8:192-200); the Box P-binding protein (the BPF-1) family (da Costa e Silva et al. (1993) *Plant J.* 4:125-135); the high mobility group (HMG) family (Bustin and Reeves (1996) *Prog. Nucl. Acids Res. Mol. Biol.* 54:35-100); the scarecrow (SCR) family (Di Laurenzio et al. (1996) *Cell* 86:423-433); the GF14 family (Wu et al. (1997) *Plant Physiol.* 114:1421-1431); the polycomb (PCOMB) family (Kennison (1995) *Annu. Rev. Genet.* 29:289-303); the teosinte branched (TEO) family (Luo et al. (1996) *Nature* 383:794-799); the ABI3 family (Giraudat et al. (1992) *Plant Cell* 4:1251-1261); the triple helix (TH) family (Dehesh et al. (1990) *Science* 250:1397-1399); the EIL family (Chao et al. (1997) *Cell* 89:1133-44); the AT-HOOK family (Reeves and Nissen (1990) *J. Biol. Chem.* 265:8573-8582); the S1FA family (Zhou et al. (1995) *Nucleic Acids Res.* 23:1165-1169); the bZIPT2 family (Lu and Ferl (1995) *Plant Physiol.* 109:723); the YABBY family (Bowman et al. (1999) *Development* 126:2387-96); the PAZ family (Bohmert et al. (1998) *EMBO J.* 17:170-80); a family of miscellaneous (MISC) transcription factors including the DPBF family (Kim et al. (1997) *Plant J.* 11:1237-1251) and the SPF1 family (Ishiguro and Nakamura (1994) *Mol. Gen. Genet.* 244:563-571); the golden (GLD) family (Hall et al. (1998) *Plant Cell* 10:925-936), the TUBBY family (Boggin et al, (1999) *Science* 286:2119-2125), the heat shock family (Wu C (1995) *Annu Rev Cell Dev Biol* 11:441-469), the ENBP family (Christiansen et al (1996) *Plant Mol Biol* 32:809-821), the RING-zinc family (Jensen et al. (1998) *FEBS letters* 436:283-287), the PDBP family (Janik et al *Virology.* (1989) 168:320-329), the PCF family (Cubas P, et al. *Plant J.* (1999) 18:215-22), the SRS (SHI-related) family (Fridborg et al *Plant Cell* (1999) 11:1019-1032), the CPP (cysteine-rich polycomb-like) family (Cvitanich et al *Proc. Natl. Acad. Sci. U S A.* (2000) 97:8163-8168), the ARF (auxin response factor) family (Ulmasov, et al. (1999) *Proc. Natl. Acad. Sci. USA* 96: 5844-5849), the SWI/SNF family (Collingwood et al *J. Mol. End.* 23:255-275), the ACBF family (Seguin et al (1997) *Plant Mol Biol.* 35:281-291), PCGL (CG-1 like) family (da Costa e Silva et al.

(1994) *Plant Mol Biol.* 25:921-924) the ARID family (Vazquez et al. (1999) *Development.* 126: 733-42), the Jumonji family, Balciunas et al (2000, *Trends Biochem Sci.* 25: 274-276), the bZIP-NIN family (Schauser et al (1999) *Nature* 402: 191-195), the E2F family Kaelin et al (1992) *Cell* 70: 351-364) and the GRF-like family (Knaap et al (2000) *Plant Physiol.* 122: 695-704). As indicated by any part of the list above and as known in the art, transcription factors have been sometimes categorized by class, family, and sub-family according to their structural content and consensus DNA-binding site motif, for example. Many of the classes and many of the families and sub-families are listed here. However, the inclusion of one sub-family and not another, or the inclusion of one family and not another, does not mean that the invention does not encompass polynucleotides or polypeptides of a certain family or sub-family. The list provided here is merely an example of the types of transcription factors and the knowledge available concerning the consensus sequences and consensus DNA-binding site motifs that help define them as known to those of skill in the art (each of the references noted above are specifically incorporated herein by reference). A transcription factor may include, but is not limited to, any polypeptide that can activate or repress transcription of a single gene or a number of genes. This polypeptide group includes, but is not limited to, DNA-binding proteins, DNA-binding protein binding proteins, protein kinases, protein phosphatases, GTP-binding proteins, and receptors, and the like.

In addition to methods for modifying a plant phenotype by employing one or more polynucleotides and polypeptides of the invention described herein, the polynucleotides and polypeptides of the invention have a variety of additional uses. These uses include their use in the recombinant production (i.e., expression) of proteins; as regulators of plant gene expression, as diagnostic probes for the presence of complementary or partially complementary nucleic acids (including for detection of natural coding nucleic acids); as substrates for further reactions, e.g., mutation reactions, PCR reactions, or the like; as substrates for cloning e.g., including digestion or ligation reactions; and for identifying exogenous or endogenous modulators of the transcription factors. A "polynucleotide" is a nucleic acid sequence comprising a plurality of polymerized nucleotides, e.g., at least about 15 consecutive polymerized nucleotides, optionally at least about 30 consecutive nucleotides, at least about 50 consecutive nucleotides. In many instances, a polynucleotide comprises a nucleotide

sequence encoding a polypeptide (or protein) or a domain or fragment thereof. Additionally, the polynucleotide may comprise a promoter, an intron, an enhancer region, a polyadenylation site, a translation initiation site, 5' or 3' untranslated regions, a reporter gene, a selectable marker, or the like. The polynucleotide can be single stranded or double stranded DNA or RNA. The polynucleotide optionally comprises modified bases or a modified backbone. The polynucleotide can be, e.g., genomic DNA or RNA, a transcript (such as an mRNA), a cDNA, a PCR product, a cloned DNA, a synthetic DNA or RNA, or the like. The polynucleotide can comprise a sequence in either sense or antisense orientations.

A "recombinant polynucleotide" is a polynucleotide that is not in its native state, e.g., the polynucleotide comprises a nucleotide sequence not found in nature, or the polynucleotide is in a context other than that in which it is naturally found, e.g., separated from nucleotide sequences with which it typically is in proximity in nature, or adjacent (or contiguous with) nucleotide sequences with which it typically is not in proximity. For example, the sequence at issue can be cloned into a vector, or otherwise recombined with one or more additional nucleic acid.

An "isolated polynucleotide" is a polynucleotide whether naturally occurring or recombinant, that is present outside the cell in which it is typically found in nature, whether purified or not. Optionally, an isolated polynucleotide is subject to one or more enrichment or purification procedures, e.g., cell lysis, extraction, centrifugation, precipitation, or the like.

A "polypeptide" is an amino acid sequence comprising a plurality of consecutive polymerized amino acid residues e.g., at least about 15 consecutive polymerized amino acid residues, optionally at least about 30 consecutive polymerized amino acid residues, at least about 50 consecutive polymerized amino acid residues. In many instances, a polypeptide comprises a polymerized amino acid residue sequence that is a transcription factor or a domain or portion or fragment thereof. Additionally, the polypeptide may comprise a localization domain, 2) an activation domain, 3) a repression domain, 4) an oligomerization domain or 5) a DNA-binding domain, or the like. The polypeptide optionally comprises modified

amino acid residues, naturally occurring amino acid residues not encoded by a codon, non-naturally occurring amino acid residues.

A "recombinant polypeptide" is a polypeptide produced by translation of a recombinant polynucleotide. A "synthetic polypeptide" is a polypeptide created by consecutive polymerization of isolated amino acid residues using methods well known in the art. An "isolated polypeptide," whether a naturally occurring or a recombinant polypeptide, is more enriched in (or out of) a cell than the polypeptide in its natural state in a wild type cell, e.g., more than about 5% enriched, more than about 10% enriched, or more than about 20%, or more than about 50%, or more, enriched, i.e., alternatively denoted: 105%, 110%, 120%, 150% or more, enriched relative to wild type standardized at 100%. Such an enrichment is not the result of a natural response of a wild type plant. Alternatively, or additionally, the isolated polypeptide is separated from other cellular components with which it is typically associated, e.g., by any of the various protein purification methods herein.

"Identity" or "similarity" refers to sequence similarity between two polynucleotide sequences or between two polypeptide sequences, with identity being a more strict comparison. The phrases "percent identity" and "% identity" refer to the percentage of sequence similarity found in a comparison of two or more polynucleotide sequences or two or more polypeptide sequences. Identity or similarity can be determined by comparing a position in each sequence that may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same nucleotide base or amino acid, then the molecules are identical at that position. A degree of similarity or identity between polynucleotide sequences is a function of the number of identical or matching nucleotides at positions shared by the polynucleotide sequences. A degree of identity of polypeptide sequences is a function of the number of identical amino acids at positions shared by the polypeptide sequences. A degree of homology or similarity of polypeptide sequences is a function of the number of amino acids, i.e., structurally related, at positions shared by the polypeptide sequences.

"Altered" nucleic acid sequences encoding polypeptide include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting

in a polynucleotide encoding a polypeptide with at least one functional characteristic of the polypeptide. Included within this definition are polymorphisms that may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding polypeptide, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding polypeptide. The encoded polypeptide protein may also be "altered", and may contain deletions, insertions, or substitutions of amino acid residues that produce a silent change and result in a functionally equivalent polypeptide. Deliberate amino acid substitutions may be made on the basis of similarity in residue side chain chemistry, including, but not limited to, polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological activity of polypeptide is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, positively charged amino acids may include lysine and arginine, and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; and phenylalanine and tyrosine. Alignments between different polypeptide sequences may be used to calculate "percentage sequence similarity".

The term "plant" includes whole plants, shoot vegetative organs/structures (e.g., leaves, stems and tubers), roots, flowers and floral organs/structures (e.g., bracts, sepals, petals, stamens, carpels, anthers and ovules), seed (including embryo, endosperm, and seed coat) and fruit (the mature ovary), plant tissue (e.g., vascular tissue, ground tissue, and the like) and cells (e.g., guard cells, egg cells, and the like), and progeny of same. The class of plants that can be used in the method of the invention is generally as broad as the class of higher and lower plants amenable to transformation techniques, including angiosperms (monocotyledonous and dicotyledonous plants), gymnosperms, ferns, horsetails, psilophytes, lycophytes, bryophytes, and multicellular algae. (See for example, Figure 1, adapted from Daly et al. 2001 *Plant Physiology* 127:1328-1333; and see also Tudge, C., The Variety of Life, Oxford University Press, New York, 2000, pp. 547-606.)

A "transgenic plant" refers to a plant that contains genetic material not found in a wild type plant of the same species, variety or cultivar. The genetic material may

include a transgene, an insertional mutagenesis event (such as by transposon or T-DNA insertional mutagenesis), an activation tagging sequence, a mutated sequence, a homologous recombination event or a sequence modified by chimeraplasty. Typically, the foreign genetic material has been introduced into the plant by human manipulation, but any method can be used as one of skill in the art recognizes.

A transgenic plant may contain an expression vector or cassette. The expression cassette typically comprises a polypeptide-encoding sequence operably linked (i.e., under regulatory control of) to appropriate inducible or constitutive regulatory sequences that allow for the expression of polypeptide. The expression cassette can be introduced into a plant by transformation or by breeding after transformation of a parent plant. A plant refers to a whole plant as well as to a plant part, such as seed, fruit, leaf, or root, plant tissue, plant cells or any other plant material, e.g., a plant explant, as well as to progeny thereof, and to *in vitro* systems that mimic biochemical or cellular components or processes in a cell.

"Ectopic expression or altered expression" in reference to a polynucleotide indicates that the pattern of expression in, e.g., a transgenic plant or plant tissue, is different from the expression pattern in a wild type plant or a reference plant of the same species. The pattern of expression may also be compared with a reference expression pattern in a wild type plant of the same species. For example, the polynucleotide or polypeptide is expressed in a cell or tissue type other than a cell or tissue type in which the sequence is expressed in the wild type plant, or by expression at a time other than at the time the sequence is expressed in the wild type plant, or by a response to different inducible agents, such as hormones or environmental signals, or at different expression levels (either higher or lower) compared with those found in a wild type plant. The term also refers to altered expression patterns that are produced by lowering the levels of expression to below the detection level or completely abolishing expression. The resulting expression pattern can be transient or stable, constitutive or inducible. In reference to a polypeptide, the term "ectopic expression or altered expression" further may relate to altered activity levels resulting from the interactions of the polypeptides with exogenous or endogenous modulators or from interactions with factors or as a result of the chemical modification of the polypeptides.

A "fragment" or "domain," with respect to a polypeptide, refers to a subsequence of the polypeptide. In some cases, the fragment or domain, is a subsequence of the polypeptide which performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA-binding site or domain that binds to a DNA promoter region, an activation domain, or a domain for protein-protein interactions. Fragments can vary in size from as few as 6 amino acids to the full length of the intact polypeptide, but are preferably at least about 30 amino acids in length and more preferably at least about 60 amino acids in length. In reference to a polynucleotide sequence, "a fragment" refers to any subsequence of a polynucleotide, typically, of at least about 15 consecutive nucleotides, preferably at least about 30 nucleotides, more preferably at least about 50 nucleotides, of any of the sequences provided herein.

The invention also encompasses production of DNA sequences that encode transcription factors and transcription factor derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding transcription factors or any fragment thereof.

A "conserved domain", with respect to a polypeptide, refers to a domain within a transcription factor family which exhibits a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% amino acid residue sequence identity of a polypeptide of consecutive amino acid residues. A fragment or domain can be referred to as outside a consensus sequence or outside a consensus DNA-binding site that is known to exist or that exists for a particular transcription factor class, family, or sub-family. In this case, the fragment or domain will not include the exact amino acids of a consensus sequence or consensus DNA-binding

site of a transcription factor class, family or sub-family, or the exact amino acids of a particular transcription factor consensus sequence or consensus DNA-binding site. Furthermore, a particular fragment, region, or domain of a polypeptide, or a polynucleotide encoding a polypeptide, can be "outside a conserved domain" if all the amino acids of the fragment, region, or domain fall outside of a defined conserved domain(s) for a polypeptide or protein. The conserved domains for each of polypeptides of SEQ ID NOs:2 - 2N, where $N = 2-561$, are listed in Table 4 as described in Example VII. Also, many of the polypeptides of Table 4 have conserved domains specifically indicated by start and stop sites. A comparison of the regions of the polypeptides in SEQ ID NOs:2 - 2N, where $N = 2-561$, or of those in Table 4, allows one of skill in the art to identify conserved domain(s) for any of the polypeptides listed or referred to in this disclosure, including those in Table 4.

A "trait" refers to a physiological, morphological, biochemical, or physical characteristic of a plant or particular plant material or cell. In some instances, this characteristic is visible to the human eye, such as seed or plant size, or can be measured by biochemical techniques, such as detecting the protein, starch, or oil content of seed or leaves, or by observation of a metabolic or physiological process, e.g. by measuring uptake of carbon dioxide, or by the observation of the expression level of a gene or genes, e.g., by employing Northern analysis, RT-PCR, microarray gene expression assays, or reporter gene expression systems, or by agricultural observations such as stress tolerance, yield, or pathogen tolerance. Any technique can be used to measure the amount of, comparative level of, or difference in any selected chemical compound or macromolecule in the transgenic plants, however.

"Trait modification" refers to a detectable difference in a characteristic in a plant ectopically expressing a polynucleotide or polypeptide of the present invention relative to a plant not doing so, such as a wild type plant. In some cases, the trait modification can be evaluated quantitatively. For example, the trait modification can entail at least about a 2% increase or decrease in an observed trait (difference); at least a 5% difference, at least about a 10% difference, at least about a 20% difference, at least about a 30%, at least about a 50%, at least about a 70%, or at least about a 100%, or an even greater difference compared with a wild type plant. It is known that there can be a natural variation in the modified trait. Therefore, the trait modification

observed entails a change of the normal distribution of the trait in the plants compared with the distribution observed in wild type plant.

I. Traits Which May Be Modified

Trait modifications of particular interest include those to seed (such as embryo or endosperm), fruit, root, flower, leaf, stem, shoot, seedling or the like, including: enhanced tolerance to environmental conditions including freezing, chilling, heat, drought, water saturation, radiation and ozone; improved tolerance to microbial, fungal or viral diseases; improved tolerance to pest infestations, including nematodes, mollicutes, parasitic higher plants or the like; decreased herbicide sensitivity; improved tolerance of heavy metals or enhanced ability to take up heavy metals; improved growth under poor photoconditions (e.g., low light and/or short day length), or changes in expression levels of genes of interest. Other phenotype that can be modified relate to the production of plant metabolites, such as variations in the production of taxol, tocopherol, tocotrienol, sterols, phytosterols, vitamins, wax monomers, anti-oxidants, amino acids, lignins, cellulose, tannins, prenolipids (such as chlorophylls and carotenoids), glucosinolates, and terpenoids, enhanced or compositionally altered protein or oil production (especially in seeds), or modified sugar (insoluble or soluble) and/or starch composition. Physical plant characteristics that can be modified include cell development (such as the number of trichomes), fruit and seed size and number, yields of plant parts such as stems, leaves, inflorescences, and roots, the stability of the seeds during storage, characteristics of the seed pod (e.g., susceptibility to shattering), root hair length and quantity, internode distances, or the quality of seed coat. Plant growth characteristics that can be modified include growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, as well as plant architecture characteristics such as apical dominance, branching patterns, number of organs, organ identity, organ shape or size.

II. Transcription Factors Modify Expression Of Endogenous Genes

Expression of genes which encode transcription factors that modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription

factors may also modify expression of endogenous genes, polynucleotides, and proteins. Examples include Peng et al. (1997, *Genes and Development* 11:3194-3205) and Peng et al. (1999, *Nature*, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, *Plant Cell* 13:1791-1802); Nandi et al. (2000, *Curr. Biol.* 10:215-218); Coupland (1995, *Nature* 377:482-483); and Weigel and Nilsson (1995, *Nature* 377:482-500).

In another example, Mandel et al. (1992, *Cell* 71:133-143) and Suzuki et al. (2001, *Plant J.* 28:409-418) teach that a transcription factor expressed in another plant species elicits the same or very similar phenotypic response of the endogenous sequence, as often predicted in earlier studies of *Arabidopsis* transcription factors in *Arabidopsis* (see Mandel et al., 1992, *supra*; Suzuki et al., 2001, *supra*).

Other examples include Müller et al. (2001, *Plant J.* 28:169-179); Kim et al. (2001, *Plant J.* 25:247-259); Kyoizuka and Shimamoto (2002, *Plant Cell Physiol.* 43:130-135); Boss and Thomas (2002, *Nature*, 416:847-850); He et al. (2000, *Transgenic Res.*, 9:223-227); and Robson et al. (2001, *Plant J.* 28:619-631).

In yet another example, Gilmour et al. (1998, *Plant J.* 16:433-442) teach an *Arabidopsis* AP2 transcription factor, CBF1, which, when overexpressed in transgenic plants, increases plant freezing tolerance. Jaglo et al (2001, *Plant Physiol.* 127:910-917) further identified sequences in *Brassica napus* which encode CBF-like genes and that transcripts for these genes accumulated rapidly in response to low temperature. Transcripts encoding CBF-like proteins were also found to accumulate rapidly in response to low temperature in wheat, as well as in tomato. An alignment of the CBF proteins from *Arabidopsis*, *B. napus*, wheat, rye, and tomato revealed the presence of conserved amino acid sequences, PKK/RPAGR_xKFxETRHP and DSAWR, that bracket the AP2/EREBP DNA binding domains of the proteins and distinguish them from other members of the AP2/EREBP protein family. (See Jaglo et al., *supra*.)

III. Polypeptides and Polynucleotides of the Invention

The present invention provides, among other things, transcription factors (TFs), and transcription factor homologue polypeptides, and isolated or recombinant polynucleotides encoding the polypeptides, or novel variant polypeptides or polynucleotides encoding novel variants of transcription factors derived from the specific sequences provided here. These polypeptides and polynucleotides may be employed to modify a plant's characteristic.

Exemplary polynucleotides encoding the polypeptides of the invention were identified in the *Arabidopsis thaliana* GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. In addition, further exemplary polynucleotides encoding the polypeptides of the invention were identified in the plant GenBank database using publicly available sequence analysis programs and parameters. Sequences initially identified were then further characterized to identify sequences comprising specified sequence strings corresponding to sequence motifs present in families of known transcription factors. Polynucleotide sequences meeting such criteria were confirmed as transcription factors.

Additional polynucleotides of the invention were identified by screening *Arabidopsis thaliana* and/or other plant cDNA libraries with probes corresponding to known transcription factors under low stringency hybridization conditions. Additional sequences, including full length coding sequences were subsequently recovered by the rapid amplification of cDNA ends (RACE) procedure, using a commercially available kit according to the manufacturer's instructions. Where necessary, multiple rounds of RACE are performed to isolate 5' and 3' ends. The full length cDNA was then recovered by a routine end-to-end polymerase chain reaction (PCR) using primers specific to the isolated 5' and 3' ends. Exemplary sequences are provided in the Sequence Listing.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor or knockout plants and the changes in the characteristic(s) or trait(s) of

the plants observed. Therefore, the polynucleotides and polypeptides can be employed to improve the characteristics of plants.

The polynucleotides of the invention can be or were ectopically expressed in overexpressor plant cells and the changes in the expression levels of a number of genes, polynucleotides, and/or proteins of the plant cells observed. Therefore, the polynucleotides and polypeptides can be employed to change expression levels of a genes, polynucleotides, and/or proteins of plants.

IV. Producing Polypeptides

The polynucleotides of the invention include sequences that encode transcription factors and transcription factor homologue polypeptides and sequences complementary thereto, as well as unique fragments of coding sequence, or sequence complementary thereto. Such polynucleotides can be, e.g., DNA or RNA, e.g., mRNA, cRNA, synthetic RNA, genomic DNA, cDNA synthetic DNA, oligonucleotides, etc. The polynucleotides are either double-stranded or single-stranded, and include either, or both sense (i.e., coding) sequences and antisense (i.e., non-coding, complementary) sequences. The polynucleotides include the coding sequence of a transcription factor, or transcription factor homologue polypeptide, in isolation, in combination with additional coding sequences (e.g., a purification tag, a localization signal, as a fusion-protein, as a pre-protein, or the like), in combination with non-coding sequences (e.g., introns or inteins, regulatory elements such as promoters, enhancers, terminators, and the like), and/or in a vector or host environment in which the polynucleotide encoding a transcription factor or transcription factor homologue polypeptide is an endogenous or exogenous gene.

A variety of methods exist for producing the polynucleotides of the invention. Procedures for identifying and isolating DNA clones are well known to those of skill in the art, and are described in, e.g., Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology volume 152 Academic Press, Inc., San Diego, CA ("Berger"); Sambrook et al., Molecular Cloning - A Laboratory Manual (2nd Ed.), Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York, 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds.,

Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 2000) ("Ausubel").

Alternatively, polynucleotides of the invention, can be produced by a variety of in vitro amplification methods adapted to the present invention by appropriate selection of specific or degenerate primers. Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR) the ligase chain reaction (LCR), Qbeta-replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger (supra), Sambrook (supra), and Ausubel (supra), as well as Mullis et al., (1987) PCR Protocols A Guide to Methods and Applications (Innis et al. eds) Academic Press Inc. San Diego, CA (1990) (Innis). Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No. 5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, e.g., Ausubel, Sambrook and Berger, *all supra*.

Alternatively, polynucleotides and oligonucleotides of the invention can be assembled from fragments produced by solid-phase synthesis methods. Typically, fragments of up to approximately 100 bases are individually synthesized and then enzymatically or chemically ligated to produce a desired sequence, e.g., a polynucleotide encoding all or part of a transcription factor. For example, chemical synthesis using the phosphoramidite method is described, e.g., by Beaucage et al. (1981) Tetrahedron Letters 22:1859-1869; and Matthes et al. (1984) EMBO J. 3:801-805. According to such methods, oligonucleotides are synthesized, purified, annealed to their complementary strand, ligated and then optionally cloned into suitable vectors. And if so desired, the polynucleotides and polypeptides of the invention can be custom ordered from any of a number of commercial suppliers.

V. Homologous Sequences

Sequences homologous, i.e., that share significant sequence identity or similarity, to those provided in the Sequence Listing, derived from *Arabidopsis thaliana* or from other plants of choice are also an aspect of the invention.

Homologous sequences can be derived from any plant including monocots and dicots and in particular agriculturally important plant species, including but not limited to, crops such as soybean, wheat, corn, potato, cotton, rice, rape, oilseed rape (including canola), sunflower, alfalfa, sugarcane and turf; or fruits and vegetables, such as banana, blackberry, blueberry, strawberry, and raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, rosaceous fruits (such as apple, peach, pear, cherry and plum) and vegetable brassicas (such as broccoli, cabbage, cauliflower, Brussels sprouts, and kohlrabi). Other crops, fruits and vegetables whose phenotype can be changed include barley, rye, millet, sorghum, currant, avocado, citrus fruits such as oranges, lemons, grapefruit and tangerines, artichoke, cherries, nuts such as the walnut and peanut, endive, leek, roots, such as arrowroot, beet, cassava, turnip, radish, yam, and sweet potato, and beans. The homologous sequences may also be derived from woody species, such as pine, poplar and eucalyptus, or mint or other labiates.

Orthologs And Paralogs

Several different methods are known by those of skill in the art for identifying and defining these functionally homologous sequences. Three general methods for defining paralogs and orthologs are described; a paralog or ortholog or homolog may be identified by one or more of the methods described below.

Orthologs and paralogs are evolutionarily related genes that have similar sequence and similar functions. Orthologs are structurally related genes in different species that are derived from a speciation event. Paralogs are structurally related genes within a single species that are derived by a duplication event.

Within a single plant species, gene duplication may cause two copies of a particular gene, giving rise to two or more genes with similar sequence and similar function known as paralogs. A paralog is therefore a similar gene with a similar function within the same species. Paralogs typically cluster together or in the same

clade (a group of similar genes) when a gene family phylogeny is analyzed using programs such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266 383-402). Groups of similar genes can also be identified with pair-wise BLAST analysis (Feng and Doolittle (1987) *J. Mol. Evol.* 25:351-360). For example, a clade of very similar MADS domain transcription factors from *Arabidopsis* all share a common function in flowering time (Ratcliffe et al. (2001) *Plant Physiol.* 126:122-132), and a group of very similar AP2 domain transcription factors from *Arabidopsis* are involved in tolerance of plants to freezing (Gilmour et al. (1998) *Plant J.* 16:433-442). Analysis of groups of similar genes with similar function that fall within one clade can yield sub-sequences that are particular to the clade. These sub-sequences, known as consensus sequences, can not only be used to define the sequences within each clade, but define the functions of these genes; genes within a clade may contain paralogous or orthologous sequences that share the same function. (See also, for example, Mount, D.W. (2001) Bioinformatics: Sequence and Genome Analysis Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York page 543.)

Speciation, the production of new species from a parental species, can also give rise to two or more genes with similar sequence and similar function. These genes, termed orthologs, often have an identical function within their host plants and are often interchangeable between species without losing function. Because plants have common ancestors, many genes in any plant species will have a corresponding orthologous gene in another plant species. Once a phylogenetic tree for a gene family of one species has been constructed using a program such as CLUSTAL (Thompson et al. (1994) *Nucleic Acids Res.* 22:4673-4680; Higgins et al. (1996) *Methods Enzymol.* 266:383-402), potential orthologous sequences can be placed into the phylogenetic tree and its relationship to genes from the species of interest can be determined. Once the ortholog pair has been identified, the function of the test ortholog can be determined by determining the function of the reference ortholog.

Transcription factors that are homologous to the listed sequences will typically share at least about 30% amino acid sequence identity, or at least about 30% amino acid sequence identity outside of a known consensus sequence or consensus DNA-binding site. More closely related transcription factors can share at least about 50%,

about 60%, about 65%, about 70%, about 75% or about 80% or about 90% or about 95% or about 98% or more sequence identity with the listed sequences, or with the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site, or with the listed sequences excluding one or all conserved domain. Factors that are most closely related to the listed sequences share, e.g., at least about 85%, about 90% or about 95% or more % sequence identity to the listed sequences, or to the listed sequences but excluding or outside a known consensus sequence or consensus DNA-binding site or outside one or all conserved domain. At the nucleotide level, the sequences will typically share at least about 40% nucleotide sequence identity, preferably at least about 50%, about 60%, about 70% or about 80% sequence identity, and more preferably about 85%, about 90%, about 95% or about 97% or more sequence identity to one or more of the listed sequences, or to a listed sequence but excluding or outside a known consensus sequence or consensus DNA-binding site, or outside one or all conserved domain. The degeneracy of the genetic code enables major variations in the nucleotide sequence of a polynucleotide while maintaining the amino acid sequence of the encoded protein. Conserved domains within a transcription factor family may exhibit a higher degree of sequence homology, such as at least 65% sequence identity including conservative substitutions, and preferably at least 80% sequence identity, and more preferably at least 85%, or at least about 86%, or at least about 87%, or at least about 88%, or at least about 90%, or at least about 95%, or at least about 98% sequence identity. Transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence identity over the entire length of the polypeptide or the homolog. In addition, transcription factors that are homologous to the listed sequences should share at least 30%, or at least about 60%, or at least about 75%, or at least about 80%, or at least about 90%, or at least about 95% amino acid sequence similarity over the entire length of the polypeptide or the homolog.

Percent identity can be determined electronically, e.g., by using the MEGALIGN program (DNASTAR, Inc. Madison, Wis.). The MEGALIGN program can create alignments between two or more sequences according to different methods, e.g., the clustal method. (See, e.g., Higgins, D. G. and P. M. Sharp (1988) Gene

73:237-244.) The clustal algorithm groups sequences into clusters by examining the distances between all pairs. The clusters are aligned pairwise and then in groups. Other alignment algorithms or programs may be used, including FASTA, BLAST, or ENTREZ, FASTA and BLAST. These are available as a part of the GCG sequence analysis package (University of Wisconsin, Madison, Wis.), and can be used with or without default settings. ENTREZ is available through the National Center for Biotechnology Information. In one embodiment, the percent identity of two sequences can be determined by the GCG program with a gap weight of 1, e.g., each amino acid gap is weighted as if it were a single amino acid or nucleotide mismatch between the two sequences (see USPN 6,262,333).

Other techniques for alignment are described in *Methods in Enzymology*, vol. 266: *Computer Methods for Macromolecular Sequence Analysis* (1996), ed. Doolittle, Academic Press, Inc., San Diego, Calif., USA. Preferably, an alignment program that permits gaps in the sequence is utilized to align the sequences. The Smith-Waterman is one type of algorithm that permits gaps in sequence alignments. See *Methods Mol. Biol.* 70: 173-187 (1997). Also, the GAP program using the Needleman and Wunsch alignment method can be utilized to align sequences. An alternative search strategy uses MPSRCH software, which runs on a MASPAR computer. MPSRCH uses a Smith-Waterman algorithm to score sequences on a massively parallel computer. This approach improves ability to pick up distantly related matches, and is especially tolerant of small gaps and nucleotide sequence errors. Nucleic acid-encoded amino acid sequences can be used to search both protein and DNA databases.

The percentage similarity between two polypeptide sequences, e.g., sequence A and sequence B, is calculated by dividing the length of sequence A, minus the number of gap residues in sequence A, minus the number of gap residues in sequence B, into the sum of the residue matches between sequence A and sequence B, times one hundred. Gaps of low or of no similarity between the two amino acid sequences are not included in determining percentage similarity. Percent identity between polynucleotide sequences can also be counted or calculated by other methods known in the art, e.g., the Jotun Hein method. (See, e.g., Hein, J. (1990) *Methods Enzymol.* 183:626-645.) Identity between sequences can also be determined by other methods

known in the art, e.g., by varying hybridization conditions (see US Patent Application No. 20010010913).

Thus, the invention provides methods for identifying a sequence similar or paralogous or orthologous or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

In addition, one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to search against a BLOCKS (Bairoch et al. (1997) *Nucleic Acids Res.* 25:217-221), PFAM, and other databases which contain previously identified and annotated motifs, sequences and gene functions. Methods that search for primary sequence patterns with secondary structure gap penalties (Smith et al. (1992) *Protein Engineering* 5:35-51) as well as algorithms such as Basic Local Alignment Search Tool (BLAST; Altschul, S. F. (1993) *J. Mol. Evol.* 36:290-300; Altschul et al. (1990) *supra*), BLOCKS (Henikoff, S. and Henikoff, G. J. (1991) *Nucleic Acids Research* 19:6565-6572), Hidden Markov Models (HMM; Eddy, S. R. (1996) *Cur. Opin. Str. Biol.* 6:361-365; Sonnhammer et al. (1997) *Proteins* 28:405-420), and the like, can be used to manipulate and analyze polynucleotide and polypeptide sequences encoded by polynucleotides. These databases, algorithms and other methods are well known in the art and are described in Ausubel et al. (1997; *Short Protocols in Molecular Biology*, John Wiley & Sons, New York N.Y., unit 7.7) and in Meyers, R. A. (1995; *Molecular Biology and Biotechnology*, Wiley VCH, New York N.Y., p 856-853).

Furthermore, methods using manual alignment of sequences similar or homologous to one or more polynucleotide sequences or one or more polypeptides encoded by the polynucleotide sequences may be used to identify regions of similarity and conserved domains. Such manual methods are well-known of those of skill in the art and can include, for example, comparisons of tertiary structure between a

polypeptide sequence encoded by a polynucleotide which comprises a known function with a polypeptide sequence encoded by a polynucleotide sequence which has a function not yet determined. Such examples of tertiary structure may comprise predicted alpha helices, beta-sheets, amphipathic helices, leucine zipper motifs, zinc finger motifs, proline-rich regions, cysteine repeat motifs, and the like.

VI. Identifying Polynucleotides or Nucleic Acids by Hybridization

Polynucleotides homologous to the sequences illustrated in the Sequence Listing and tables can be identified, e.g., by hybridization to each other under stringent or under highly stringent conditions. Single stranded polynucleotides hybridize when they associate based on a variety of well characterized physical-chemical forces, such as hydrogen bonding, solvent exclusion, base stacking and the like. The stringency of a hybridization reflects the degree of sequence identity of the nucleic acids involved, such that the higher the stringency, the more similar are the two polynucleotide strands. Stringency is influenced by a variety of factors, including temperature, salt concentration and composition, organic and non-organic additives, solvents, etc. present in both the hybridization and wash solutions and incubations (and number thereof), as described in more detail in the references cited above. Encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed polynucleotide sequences, and, in particular, to those shown in SEQ ID NOs: 860; 802; 240; 274; 558; 24; 1120; 44; 460; 286; 120; 130; 134; 698; 832; 580; 612; 48, and fragments thereof under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399-407; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507-511.) Estimates of homology are provided by either DNA-DNA or DNA-RNA hybridization under conditions of stringency as is well understood by those skilled in the art (Hames and Higgins, Eds. (1985) *Nucleic Acid Hybridisation*, IRL Press, Oxford, U.K.). Stringency conditions can be adjusted to screen for moderately similar fragments, such as homologous sequences from distantly related organisms, to highly similar fragments, such as genes that duplicate functional enzymes from closely related organisms. Post-hybridization washes determine stringency conditions.

In addition to the nucleotide sequences listed in Tables 4 and 5, full length cDNA, orthologs, paralogs and homologs of the present nucleotide sequences may be

identified and isolated using well known methods. The cDNA libraries orthologs, paralogs and homologs of the present nucleotide sequences may be screened using hybridization methods to determine their utility as hybridization target or amplification probes.

An example of stringent hybridization conditions for hybridization of complementary nucleic acids which have more than 100 complementary residues on a filter in a Southern or northern blot is about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. Nucleic acid molecules that hybridize under stringent conditions will typically hybridize to a probe based on either the entire cDNA or selected portions, e.g., to a unique subsequence, of the cDNA under wash conditions of 0.2x SSC to 2.0 x SSC, 0.1% SDS at 50-65° C. For example, high stringency is about 0.2 x SSC, 0.1% SDS at 65° C. Ultra-high stringency will be the same conditions except the wash temperature is raised about 3 to about 5° C, and ultra-ultra-high stringency will be the same conditions except the wash temperature is raised about 6 to about 9° C. For identification of less closely related homologues washes can be performed at a lower temperature, e.g., 50° C. In general, stringency is increased by raising the wash temperature and/or decreasing the concentration of SSC, as known in the art.

In another example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C, more preferably of at least about 37° C, and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the

art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

The washing steps that follow hybridization can also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include temperature of at least about 25° C, more preferably of at least about 42° C. Another preferred set of highly stringent conditions uses two final washes in 0.1X SSC, 0.1% SDS at 65° C. The most preferred high stringency washes are of at least about 68° C. For example, in a preferred embodiment, wash steps will occur at 25° C in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, the wash steps will occur at 68° C in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional variations on these conditions will be readily apparent to those skilled in the art (see U.S. Patent Application No. 20010010913).

As another example, stringent conditions can be selected such that an oligonucleotide that is perfectly complementary to the coding oligonucleotide hybridizes to the coding oligonucleotide with at least about a 5-10x higher signal to noise ratio than the ratio for hybridization of the perfectly complementary oligonucleotide to a nucleic acid encoding a transcription factor known as of the filing

date of the application. Conditions can be selected such that a higher signal to noise ratio is observed in the particular assay which is used, e.g., about 15x, 25x, 35x, 50x or more. Accordingly, the subject nucleic acid hybridizes to the unique coding oligonucleotide with at least a 2x higher signal to noise ratio as compared to hybridization of the coding oligonucleotide to a nucleic acid encoding known polypeptide. Again, higher signal to noise ratios can be selected, e.g., about 5x, 10x, 25x, 35x, 50x or more. The particular signal will depend on the label used in the relevant assay, e.g., a fluorescent label, a colorimetric label, a radioactive label, or the like.

Alternatively, transcription factor homolog polypeptides can be obtained by screening an expression library using antibodies specific for one or more transcription factors. With the provision herein of the disclosed transcription factor, and transcription factor homologue nucleic acid sequences, the encoded polypeptide(s) can be expressed and purified in a heterologous expression system (e.g., *E. coli*) and used to raise antibodies (monoclonal or polyclonal) specific for the polypeptide(s) in question. Antibodies can also be raised against synthetic peptides derived from transcription factor, or transcription factor homologue, amino acid sequences. Methods of raising antibodies are well known in the art and are described in Harlow and Lane (1988) Antibodies: A Laboratory Manual, Cold Spring Harbor Laboratory, New York. Such antibodies can then be used to screen an expression library produced from the plant from which it is desired to clone additional transcription factor homologues, using the methods described above. The selected cDNAs can be confirmed by sequencing and enzymatic activity.

VII. Sequence Variations

It will readily be appreciated by those of skill in the art, that any of a variety of polynucleotide sequences are capable of encoding the transcription factors and transcription factor homologue polypeptides of the invention. Due to the degeneracy of the genetic code, many different polynucleotides can encode identical and/or substantially similar polypeptides in addition to those sequences illustrated in the Sequence Listing. Nucleic acids having a sequence that differs from the sequences shown in the Sequence Listing, or complementary sequences, that encode functionally equivalent peptides (i.e., peptides having some degree of equivalent or similar

biological activity) but differ in sequence from the sequence shown in the sequence listing due to degeneracy in the genetic code, are also within the scope of the invention.

Altered polynucleotide sequences encoding polypeptides include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polynucleotide encoding a polypeptide with at least one functional characteristic of the instant polypeptides. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding the instant polypeptides, and improper or unexpected hybridization to allelic variants, with a locus other than the normal chromosomal locus for the polynucleotide sequence encoding the instant polypeptides.

Allelic variant refers to any of two or more alternative forms of a gene occupying the same chromosomal locus. Allelic variation arises naturally through mutation, and may result in phenotypic polymorphism within populations. Gene mutations can be silent (i.e., no change in the encoded polypeptide) or may encode polypeptides having altered amino acid sequence. The term allelic variant is also used herein to denote a protein encoded by an allelic variant of a gene. Splice variant refers to alternative forms of RNA transcribed from a gene. Splice variation arises naturally through use of alternative splicing sites within a transcribed RNA molecule, or less commonly between separately transcribed RNA molecules, and may result in several mRNAs transcribed from the same gene. Splice variants may encode polypeptides having altered amino acid sequence. The term splice variant is also used herein to denote a protein encoded by a splice variant of an mRNA transcribed from a gene.

Those skilled in the art would recognize that the polypeptide sequence G681, SEQ ID NO: 580, represents a single transcription factor; allelic variation and alternative splicing may be expected to occur. Allelic variants of the polypeptide sequence of SEQ ID NO: 579 can be cloned by probing cDNA or genomic libraries from different individual organisms according to standard procedures. Allelic variants of the DNA sequence shown in SEQ ID NO: 579, including those containing silent mutations and those in which mutations result in amino acid sequence changes, are within the scope of the present invention, as are proteins which are allelic variants

of SEQ ID NO: 580. cDNAs generated from alternatively spliced mRNAs, which retain the properties of the transcription factor are included within the scope of the present invention, as are polypeptides encoded by such cDNAs and mRNAs. Allelic variants and splice variants of these sequences can be cloned by probing cDNA or genomic libraries from different individual organisms or tissues according to standard procedures known in the art (see USPN 6,388,064).

For example, Table 1 illustrates, e.g., that the codons AGC, AGT, TCA, TCC, TCG, and TCT all encode the same amino acid: serine. Accordingly, at each position in the sequence where there is a codon encoding serine, any of the above trinucleotide sequences can be used without altering the encoded polypeptide.

Table 1

Amino acid			Possible Codons							
Alanine	Ala	A	GCA	GCC	GCG	GCU				
Cysteine	Cys	C	TGC	TGT						
Aspartic acid	Asp	D	GAC	GAT						
Glutamic acid	Glu	E	GAA	GAG						
Phenylalanine	Phe	F	TTC	TTT						
Glycine	Gly	G	GGA	GGC	GGG	GGT				
Histidine	His	H	CAC	CAT						
Isoleucine	Ile	I	ATA	ATC	ATT					
Lysine	Lys	K	AAA	AAG						
Leucine	Leu	L	TTA	TTG	CTA	CTC	CTG	CTT		
Methionine	Met	M	ATG							
Asparagine	Asn	N	AAC	AAT						
Proline	Pro	P	CCA	CCC	CCG	CCT				
Glutamine	Gln	Q	CAA	CAG						
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGT		
Serine	Ser	S	AGC	AGT	TCA	TCC	TCG	TCT		
Threonine	Thr	T	ACA	ACC	ACG	ACT				
Valine	Val	V	GTA	GTC	GTG	GTT				
Tryptophan	Trp	W	TGG							
Tyrosine	Tyr	Y	TAC	TAT						

Sequence alterations that do not change the amino acid sequence encoded by the polynucleotide are termed "silent" variations. With the exception of the codons ATG and TGG, encoding methionine and tryptophan, respectively, any of the possible codons for the same amino acid can be substituted by a variety of techniques, e.g., site-directed mutagenesis, available in the art. Accordingly, any and all such variations of a sequence selected from the above table are a feature of the invention.

In addition to silent variations, other conservative variations that alter one, or a few amino acids in the encoded polypeptide, can be made without altering the function of the polypeptide, these conservative variants are, likewise, a feature of the invention.

For example, substitutions, deletions and insertions introduced into the sequences provided in the Sequence Listing are also envisioned by the invention. Such sequence modifications can be engineered into a sequence by site-directed mutagenesis (Wu (ed.) Meth. Enzymol. (1993) vol. 217, Academic Press) or the other methods noted below. Amino acid substitutions are typically of single residues; insertions usually will be on the order of about from 1 to 10 amino acid residues; and deletions will range about from 1 to 30 residues. In preferred embodiments, deletions or insertions are made in adjacent pairs, e.g., a deletion of two residues or insertion of two residues. Substitutions, deletions, insertions or any combination thereof can be combined to arrive at a sequence. The mutations that are made in the polynucleotide encoding the transcription factor should not place the sequence out of reading frame and should not create complementary regions that could produce secondary mRNA structure. Preferably, the polypeptide encoded by the DNA performs the desired function.

Conservative substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 2 when it is desired to maintain the activity of the protein. Table 2 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as conservative substitutions.

Table 2

Residue	Conservative Substitutions
Ala	Ser
Arg	Lys
Asn	Gln; His
Asp	Glu
Gln	Asn
Cys	Ser
Glu	Asp
Gly	Pro
His	Asn; Gln
Ile	Leu, Val
Leu	Ile; Val
Lys	Arg; Gln
Met	Leu; Ile
Phe	Met; Leu; Tyr
Ser	Thr; Gly
Thr	Ser; Val
Trp	Tyr
Tyr	Trp; Phe
Val	Ile; Leu

Similar substitutions are those in which at least one residue in the amino acid sequence has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the Table 3 when it is desired to maintain the activity of the protein. Table 3 shows amino acids which can be substituted for an amino acid in a protein and which are typically regarded as structural and functional substitutions. For example, a residue in column 1 of Table 3

may be substituted with residue in column 2; in addition, a residue in column 2 of Table 3 may be substituted with the residue of column 1.

Table 3

Residue	Similar Substitutions
Ala	Ser; Thr; Gly; Val; Leu; Ile
Arg	Lys; His; Gly
Asn	Gln; His; Gly; Ser; Thr
Asp	Glu; Ser; Thr
Gln	Asn; Ala
Cys	Ser; Gly
Glu	Asp
Gly	Pro; Arg
His	Asn; Gln; Tyr; Phe; Lys; Arg
Ile	Ala; Leu; Val; Gly; Met
Leu	Ala; Ile; Val; Gly; Met
Lys	Arg; His; Gln; Gly; Pro
Met	Leu; Ile; Phe
Phe	Met; Leu; Tyr; Trp; His; Val; Ala
Ser	Thr; Gly; Asp; Ala; Val; Ile; His
Thr	Ser; Val; Ala; Gly
Trp	Tyr; Phe; His
Tyr	Trp; Phe; His
Val	Ala; Ile; Leu; Gly; Thr; Ser; Glu

Substitutions that are less conservative than those in Table 2 can be selected by picking residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site, or (c) the bulk of the side chain. The substitutions which in general are

expected to produce the greatest changes in protein properties will be those in which (a) a hydrophilic residue, e.g., seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g., leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

VIII. Further Modifying Sequences of the Invention – Mutation/Forced Evolution

In addition to generating silent or conservative substitutions as noted, above, the present invention optionally includes methods of modifying the sequences of the Sequence Listing. In the methods, nucleic acid or protein modification methods are used to alter the given sequences to produce new sequences and/or to chemically or enzymatically modify given sequences to change the properties of the nucleic acids or proteins.

Thus, in one embodiment, given nucleic acid sequences are modified, e.g., according to standard mutagenesis or artificial evolution methods to produce modified sequences. The modified sequences may be created using purified natural polynucleotides isolated from any organism or may be synthesized from purified compositions and chemicals using chemical means well known to those of skill in the art. For example, Ausubel, *supra*, provides additional details on mutagenesis methods. Artificial forced evolution methods are described, for example, by Stemmer (1994) Nature 370:389-391, Stemmer (1994) Proc. Natl. Acad. Sci. USA 91:10747-10751, and U.S. Patents 5,811,238, 5,837,500, and 6,242,568. Methods for engineering synthetic transcription factors and other polypeptides are described, for example, by Zhang et al. (2000) J. Biol. Chem. 275:33850-33860, Liu et al. (2001) J. Biol. Chem. 276:11323-11334, and Isalan et al. (2001) Nature Biotechnol. 19:656-660. Many other mutation and evolution methods are also available and expected to be within the skill of the practitioner.

Similarly, chemical or enzymatic alteration of expressed nucleic acids and polypeptides can be performed by standard methods. For example, sequence can be modified by addition of lipids, sugars, peptides, organic or inorganic compounds, by the inclusion of modified nucleotides or amino acids, or the like. For example, protein modification techniques are illustrated in Ausubel, *supra*. Further details on chemical and enzymatic modifications can be found herein. These modification methods can be used to modify any given sequence, or to modify any sequence produced by the various mutation and artificial evolution modification methods noted herein.

Accordingly, the invention provides for modification of any given nucleic acid by mutation, evolution, chemical or enzymatic modification, or other available methods, as well as for the products produced by practicing such methods, e.g., using the sequences herein as a starting substrate for the various modification approaches.

For example, optimized coding sequence containing codons preferred by a particular prokaryotic or eukaryotic host can be used e.g., to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced using a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *Saccharomyces cerevisiae* and mammals are TAA and TGA, respectively. The preferred stop codon for monocotyledonous plants is TGA, whereas insects and *E. coli* prefer to use TAA as the stop codon.

The polynucleotide sequences of the present invention can also be engineered in order to alter a coding sequence for a variety of reasons, including but not limited to, alterations which modify the sequence to facilitate cloning, processing and/or expression of the gene product. For example, alterations are optionally introduced using techniques which are well known in the art, e.g., site-directed mutagenesis, to insert new restriction sites, to alter glycosylation patterns, to change codon preference, to introduce splice sites, etc.

Furthermore, a fragment or domain derived from any of the polypeptides of the invention can be combined with domains derived from other transcription factors or synthetic domains to modify the biological activity of a transcription factor. For instance, a DNA-binding domain derived from a transcription factor of the invention can be combined with the activation domain of another transcription factor or with a synthetic activation domain. A transcription activation domain assists in initiating transcription from a DNA-binding site. Examples include the transcription activation region of VP16 or GAL4 (Moore et al. (1998) Proc. Natl. Acad. Sci. USA 95: 376-381; and Aoyama et al. (1995) Plant Cell 7:1773-1785), peptides derived from bacterial sequences (Ma and Ptashne (1987) Cell 51: 113-119) and synthetic peptides (Giniger and Ptashne, (1987) Nature 330:670-672).

IX. Expression and Modification of Polypeptides

Typically, polynucleotide sequences of the invention are incorporated into recombinant DNA (or RNA) molecules that direct expression of polypeptides of the invention in appropriate host cells, transgenic plants, in vitro translation systems, or the like. Due to the inherent degeneracy of the genetic code, nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can be substituted for any listed sequence to provide for cloning and expressing the relevant homologue.

X. Vectors, Promoters, and Expression Systems

The present invention includes recombinant constructs comprising one or more of the nucleic acid sequences herein. The constructs typically comprise a vector, such as a plasmid, a cosmid, a phage, a virus (e.g., a plant virus), a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

General texts that describe molecular biological techniques useful herein, including the use and production of vectors, promoters and many other relevant

topics, include Berger, Sambrook and Ausubel, *supra*. Any of the identified sequences can be incorporated into a cassette or vector, e.g., for expression in plants. A number of expression vectors suitable for stable transformation of plant cells or for the establishment of transgenic plants have been described including those described in Weissbach and Weissbach, (1989) Methods for Plant Molecular Biology, Academic Press, and Gelvin et al., (1990) Plant Molecular Biology Manual, Kluwer Academic Publishers. Specific examples include those derived from a Ti plasmid of *Agrobacterium tumefaciens*, as well as those disclosed by Herrera-Estrella et al. (1983) Nature 303: 209, Bevan (1984) Nucl Acid Res. 12: 8711-8721, Klee (1985) Bio/Technology 3: 637-642, for dicotyledonous plants.

Alternatively, non-Ti vectors can be used to transfer the DNA into monocotyledonous plants and cells by using free DNA delivery techniques. Such methods can involve, for example, the use of liposomes, electroporation, microprojectile bombardment, silicon carbide whiskers, and viruses. By using these methods transgenic plants such as wheat, rice (Christou (1991) Bio/Technology 9: 957-962) and corn (Gordon-Kamm (1990) Plant Cell 2: 603-618) can be produced. An immature embryo can also be a good target tissue for monocots for direct DNA delivery techniques by using the particle gun (Weeks et al. (1993) Plant Physiol 102: 1077-1084; Vasil (1993) Bio/Technology 10: 667-674; Wan and Lemeaux (1994) Plant Physiol 104: 37-48, and for *Agrobacterium*-mediated DNA transfer (Ishida et al. (1996) Nature Biotech 14: 745-750).

Typically, plant transformation vectors include one or more cloned plant coding sequence (genomic or cDNA) under the transcriptional control of 5' and 3' regulatory sequences and a dominant selectable marker. Such plant transformation vectors typically also contain a promoter (e.g., a regulatory region controlling inducible or constitutive, environmentally-or developmentally-regulated, or cell- or tissue-specific expression), a transcription initiation start site, an RNA processing signal (such as intron splice sites), a transcription termination site, and/or a polyadenylation signal.

Examples of constitutive plant promoters which can be useful for expressing the TF sequence include: the cauliflower mosaic virus (CaMV) 35S promoter, which

confers constitutive, high-level expression in most plant tissues (*see, e.g.,* Odell et al. (1985) Nature 313:810-812); the nopaline synthase promoter (An et al. (1988) Plant Physiol 88:547-552); and the octopine synthase promoter (Fromm et al. (1989) Plant Cell 1: 977-984).

A variety of plant gene promoters that regulate gene expression in response to environmental, hormonal, chemical, developmental signals, and in a tissue-active manner can be used for expression of a TF sequence in plants. Choice of a promoter is based largely on the phenotype of interest and is determined by such factors as tissue (e.g., seed, fruit, root, pollen, vascular tissue, flower, carpel, etc.), inducibility (e.g., in response to wounding, heat, cold, drought, light, pathogens, etc.), timing, developmental stage, and the like. Numerous known promoters have been characterized and can favorably be employed to promote expression of a polynucleotide of the invention in a transgenic plant or cell of interest. For example, tissue specific promoters include: seed-specific promoters (such as the napin, phaseolin or DC3 promoter described in US Pat. No. 5,773,697), fruit-specific promoters that are active during fruit ripening (such as the *dru 1* promoter (US Pat. No. 5,783,393), or the 2A11 promoter (US Pat. No. 4,943,674) and the tomato polygalacturonase promoter (Bird et al. (1988) Plant Mol Biol 11:651), root-specific promoters, such as those disclosed in US Patent Nos. 5,618,988, 5,837,848 and 5,905,186, pollen-active promoters such as PTA29, PTA26 and PTA13 (US Pat. No. 5,792,929), promoters active in vascular tissue (Ringli and Keller (1998) Plant Mol Biol 37:977-988), flower-specific (Kaiser et al. (1995) Plant Mol Biol 28:231-243), pollen (Baerson et al. (1994) Plant Mol Biol 26:1947-1959), carpels (Ohl et al. (1990) Plant Cell 2:837-848), pollen and ovules (Baerson et al. (1993) Plant Mol Biol 22:255-267), auxin-inducible promoters (such as that described in van der Kop et al. (1999) Plant Mol Biol 39:979-990 or Baumann et al. (1999) Plant Cell 11:323-334), cytokinin-inducible promoter (Guevara-Garcia (1998) Plant Mol Biol 38:743-753), promoters responsive to gibberellin (Shi et al. (1998) Plant Mol Biol 38:1053-1060, Willmott et al. (1998) 38:817-825) and the like. Additional promoters are those that elicit expression in response to heat (Ainley et al. (1993) Plant Mol Biol 22: 13-23), light (e.g., the pea *rbcS-3A* promoter, Kuhlemeier et al. (1989) Plant Cell 1:471, and the maize *rbcS* promoter, Schaffner and Sheen (1991) Plant Cell 3: 997); wounding (e.g., *wun1*, Siebertz et al. (1989) Plant Cell 1: 961); pathogens (such as the PR-1

promoter described in Buchel et al. (1999) Plant Mol. Biol. 40:387-396, and the PDF1.2 promoter described in Manners et al. (1998) Plant Mol. Biol. 38:1071-80), and chemicals such as methyl jasmonate or salicylic acid (Gatz et al. (1997) Plant Mol Biol 48: 89-108). In addition, the timing of the expression can be controlled by using promoters such as those acting at senescence (An and Amazon (1995) Science 270: 1986-1988); or late seed development (Odell et al. (1994) Plant Physiol 106:447-458).

Plant expression vectors can also include RNA processing signals that can be positioned within, upstream or downstream of the coding sequence. In addition, the expression vectors can include additional regulatory sequences from the 3'-untranslated region of plant genes, e.g., a 3' terminator region to increase mRNA stability of the mRNA, such as the PI-II terminator region of potato or the octopine or nopaline synthase 3' terminator regions.

Additional Expression Elements

Specific initiation signals can aid in efficient translation of coding sequences. These signals can include, e.g., the ATG initiation codon and adjacent sequences. In cases where a coding sequence, its initiation codon and upstream sequences are inserted into the appropriate expression vector, no additional translational control signals may be needed. However, in cases where only coding sequence (e.g., a mature protein coding sequence), or a portion thereof, is inserted, exogenous transcriptional control signals including the ATG initiation codon can be separately provided. The initiation codon is provided in the correct reading frame to facilitate transcription. Exogenous transcriptional elements and initiation codons can be of various origins, both natural and synthetic. The efficiency of expression can be enhanced by the inclusion of enhancers appropriate to the cell system in use.

Expression Hosts

The present invention also relates to host cells which are transduced with vectors of the invention, and the production of polypeptides of the invention (including fragments thereof) by recombinant techniques. Host cells are genetically engineered (i.e., nucleic acids are introduced, e.g., transduced, transformed or transfected) with the vectors of this invention, which may be, for example, a cloning vector or an expression vector comprising the relevant nucleic acids herein. The

vector is optionally a plasmid, a viral particle, a phage, a naked nucleic acid, *etc.* The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the relevant gene. The culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, Sambrook and Ausubel.

The host cell can be a eukaryotic cell, such as a yeast cell, or a plant cell, or the host cell can be a prokaryotic cell, such as a bacterial cell. Plant protoplasts are also suitable for some applications. For example, the DNA fragments are introduced into plant tissues, cultured plant cells or plant protoplasts by standard methods including electroporation (Fromm et al., (1985) Proc. Natl. Acad. Sci. USA 82, 5824, infection by viral vectors such as cauliflower mosaic virus (CaMV) (Hohn et al., (1982) Molecular Biology of Plant Tumors, (Academic Press, New York) pp. 549-560; US 4,407,956), high velocity ballistic penetration by small particles with the nucleic acid either within the matrix of small beads or particles, or on the surface (Klein et al., (1987) Nature 327, 70-73), use of pollen as vector (WO 85/01856), or use of *Agrobacterium tumefaciens* or *A. rhizogenes* carrying a T-DNA plasmid in which DNA fragments are cloned. The T-DNA plasmid is transmitted to plant cells upon infection by *Agrobacterium tumefaciens*, and a portion is stably integrated into the plant genome (Horsch et al. (1984) Science 233:496-498; Fraley et al. (1983) Proc. Natl. Acad. Sci. USA 80, 4803).

The cell can include a nucleic acid of the invention which encodes a polypeptide, wherein the cells expresses a polypeptide of the invention. The cell can also include vector sequences, or the like. Furthermore, cells and transgenic plants that include any polypeptide or nucleic acid above or throughout this specification, e.g., produced by transduction of a vector of the invention, are an additional feature of the invention.

For long-term, high-yield production of recombinant proteins, stable expression can be used. Host cells transformed with a nucleotide sequence encoding a polypeptide of the invention are optionally cultured under conditions suitable for the

expression and recovery of the encoded protein from cell culture. The protein or fragment thereof produced by a recombinant cell may be secreted, membrane-bound, or contained intracellularly, depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides encoding mature proteins of the invention can be designed with signal sequences which direct secretion of the mature polypeptides through a prokaryotic or eukaryotic cell membrane.

XI. Modified Amino Acid Residues

Polypeptides of the invention may contain one or more modified amino acid residues. The presence of modified amino acids may be advantageous in, for example, increasing polypeptide half-life, reducing polypeptide antigenicity or toxicity, increasing polypeptide storage stability, or the like. Amino acid residue(s) are modified, for example, co-translationally or post-translationally during recombinant production or modified by synthetic or chemical means.

Non-limiting examples of a modified amino acid residue include incorporation or other use of acetylated amino acids, glycosylated amino acids, sulfated amino acids, prenylated (e.g., farnesylated, geranylgeranylated) amino acids, PEG modified (e.g., "PEGylated") amino acids, biotinylated amino acids, carboxylated amino acids, phosphorylated amino acids, etc. References adequate to guide one of skill in the modification of amino acid residues are replete throughout the literature.

The modified amino acid residues may prevent or increase affinity of the polypeptide for another molecule, including, but not limited to, polynucleotide, proteins, carbohydrates, lipids and lipid derivatives, and other organic or synthetic compounds.

XII. Identification of Additional Factors

A transcription factor provided by the present invention can also be used to identify additional endogenous or exogenous molecules that can affect a phenotype or trait of interest. On the one hand, such molecules include organic (small or large molecules) and/or inorganic compounds that affect expression of (i.e., regulate) a particular transcription factor. Alternatively, such molecules include endogenous

molecules that are acted upon either at a transcriptional level by a transcription factor of the invention to modify a phenotype as desired. For example, the transcription factors can be employed to identify one or more downstream gene with which is subject to a regulatory effect of the transcription factor. In one approach, a transcription factor or transcription factor homologue of the invention is expressed in a host cell, e.g., a transgenic plant cell, tissue or explant, and expression products, either RNA or protein, of likely or random targets are monitored, e.g., by hybridization to a microarray of nucleic acid probes corresponding to genes expressed in a tissue or cell type of interest, by two-dimensional gel electrophoresis of protein products, or by any other method known in the art for assessing expression of gene products at the level of RNA or protein. Alternatively, a transcription factor of the invention can be used to identify promoter sequences (i.e., binding sites) involved in the regulation of a downstream target. After identifying a promoter sequence, interactions between the transcription factor and the promoter sequence can be modified by changing specific nucleotides in the promoter sequence or specific amino acids in the transcription factor that interact with the promoter sequence to alter a plant trait. Typically, transcription factor DNA-binding sites are identified by gel shift assays. After identifying the promoter regions, the promoter region sequences can be employed in double-stranded DNA arrays to identify molecules that affect the interactions of the transcription factors with their promoters (Bulyk et al. (1999) Nature Biotechnology 17:573-577).

The identified transcription factors are also useful to identify proteins that modify the activity of the transcription factor. Such modification can occur by covalent modification, such as by phosphorylation, or by protein-protein (homo or-heteropolymer) interactions. Any method suitable for detecting protein-protein interactions can be employed. Among the methods that can be employed are co-immunoprecipitation, cross-linking and co-purification through gradients or chromatographic columns, and the two-hybrid yeast system.

The two-hybrid system detects protein interactions in vivo and is described in Chien et al. ((1991), Proc. Natl. Acad. Sci. USA 88:9578-9582) and is commercially available from Clontech (Palo Alto, Calif.). In such a system, plasmids are constructed that encode two hybrid proteins: one consists of the DNA-binding domain

of a transcription activator protein fused to the TF polypeptide and the other consists of the transcription activator protein's activation domain fused to an unknown protein that is encoded by a cDNA that has been recombined into the plasmid as part of a cDNA library. The DNA-binding domain fusion plasmid and the cDNA library are transformed into a strain of the yeast *Saccharomyces cerevisiae* that contains a reporter gene (e.g., lacZ) whose regulatory region contains the transcription activator's binding site. Either hybrid protein alone cannot activate transcription of the reporter gene. Interaction of the two hybrid proteins reconstitutes the functional activator protein and results in expression of the reporter gene, which is detected by an assay for the reporter gene product. Then, the library plasmids responsible for reporter gene expression are isolated and sequenced to identify the proteins encoded by the library plasmids. After identifying proteins that interact with the transcription factors, assays for compounds that interfere with the TF protein-protein interactions can be preformed.

XIII. Identification of Modulators

In addition to the intracellular molecules described above, extracellular molecules that alter activity or expression of a transcription factor, either directly or indirectly, can be identified. For example, the methods can entail first placing a candidate molecule in contact with a plant or plant cell. The molecule can be introduced by topical administration, such as spraying or soaking of a plant, and then the molecule's effect on the expression or activity of the TF polypeptide or the expression of the polynucleotide monitored. Changes in the expression of the TF polypeptide can be monitored by use of polyclonal or monoclonal antibodies, gel electrophoresis or the like. Changes in the expression of the corresponding polynucleotide sequence can be detected by use of microarrays, Northern, quantitative PCR, or any other technique for monitoring changes in mRNA expression. These techniques are exemplified in Ausubel et al. (eds) Current Protocols in Molecular Biology, John Wiley & Sons (1998, and supplements through 2001). Such changes in the expression levels can be correlated with modified plant traits and thus identified molecules can be useful for soaking or spraying on fruit, vegetable and grain crops to modify traits in plants.

Essentially any available composition can be tested for modulatory activity of expression or activity of any nucleic acid or polypeptide herein. Thus, available libraries of compounds such as chemicals, polypeptides, nucleic acids and the like can be tested for modulatory activity. Often, potential modulator compounds can be dissolved in aqueous or organic (e.g., DMSO-based) solutions for easy delivery to the cell or plant of interest in which the activity of the modulator is to be tested. Optionally, the assays are designed to screen large modulator composition libraries by automating the assay steps and providing compounds from any convenient source to assays, which are typically run in parallel (e.g., in microtiter formats on microtiter plates in robotic assays).

In one embodiment, high throughput screening methods involve providing a combinatorial library containing a large number of potential compounds (potential modulator compounds). Such "combinatorial chemical libraries" are then screened in one or more assays, as described herein, to identify those library members (particular chemical species or subclasses) that display a desired characteristic activity. The compounds thus identified can serve as target compounds.

A combinatorial chemical library can be, e.g., a collection of diverse chemical compounds generated by chemical synthesis or biological synthesis. For example, a combinatorial chemical library such as a polypeptide library is formed by combining a set of chemical building blocks (e.g., in one example, amino acids) in every possible way for a given compound length (i.e., the number of amino acids in a polypeptide compound of a set length). Exemplary libraries include peptide libraries, nucleic acid libraries, antibody libraries (see, e.g., Vaughn et al. (1996) Nature Biotechnology, 14(3):309-314 and PCT/US96/10287), carbohydrate libraries (see, e.g., Liang et al. Science (1996) 274:1520-1522 and U.S. Patent 5,593,853), peptide nucleic acid libraries (see, e.g., U.S. Patent 5,539,083), and small organic molecule libraries (see, e.g., benzodiazepines, Baum C&EN Jan 18, page 33 (1993); isoprenoids, U.S. Patent 5,569,588; thiazolidinones and metathiazanones, U.S. Patent 5,549,974; pyrrolidines, U.S. Patents 5,525,735 and 5,519,134; morpholino compounds, U.S. Patent 5,506,337) and the like.

Preparation and screening of combinatorial or other libraries is well known to those of skill in the art. Such combinatorial chemical libraries include, but are not limited to, peptide libraries (see, e.g., U.S. Patent 5,010,175; Furka, (1991) Int. J. Pept. Prot. Res. 37:487-493; and Houghton et al. (1991) Nature 354:84-88). Other chemistries for generating chemical diversity libraries can also be used.

In addition, as noted, compound screening equipment for high-throughput screening is generally available, e.g., using any of a number of well known robotic systems that have also been developed for solution phase chemistries useful in assay systems. These systems include automated workstations including an automated synthesis apparatus and robotic systems utilizing robotic arms. Any of the above devices are suitable for use with the present invention, e.g., for high-throughput screening of potential modulators. The nature and implementation of modifications to these devices (if any) so that they can operate as discussed herein will be apparent to persons skilled in the relevant art.

Indeed, entire high throughput screening systems are commercially available. These systems typically automate entire procedures including all sample and reagent pipetting, liquid dispensing, timed incubations, and final readings of the microplate in detector(s) appropriate for the assay. These configurable systems provide high throughput and rapid start up as well as a high degree of flexibility and customization. Similarly, microfluidic implementations of screening are also commercially available.

The manufacturers of such systems provide detailed protocols the various high throughput. Thus, for example, Zymark Corp. provides technical bulletins describing screening systems for detecting the modulation of gene transcription, ligand binding, and the like. The integrated systems herein, in addition to providing for sequence alignment and, optionally, synthesis of relevant nucleic acids, can include such screening apparatus to identify modulators that have an effect on one or more polynucleotides or polypeptides according to the present invention.

In some assays it is desirable to have positive controls to ensure that the components of the assays are working properly. At least two types of positive controls are appropriate. That is, known transcriptional activators or inhibitors can be

incubated with cells/plants/ etc. in one sample of the assay, and the resulting increase/decrease in transcription can be detected by measuring the resulting increase in RNA/ protein expression, etc., according to the methods herein. It will be appreciated that modulators can also be combined with transcriptional activators or inhibitors to find modulators that inhibit transcriptional activation or transcriptional repression. Either expression of the nucleic acids and proteins herein or any additional nucleic acids or proteins activated by the nucleic acids or proteins herein, or both, can be monitored.

In an embodiment, the invention provides a method for identifying compositions that modulate the activity or expression of a polynucleotide or polypeptide of the invention. For example, a test compound, whether a small or large molecule, is placed in contact with a cell, plant (or plant tissue or explant), or composition comprising the polynucleotide or polypeptide of interest and a resulting effect on the cell, plant, (or tissue or explant) or composition is evaluated by monitoring, either directly or indirectly, one or more of: expression level of the polynucleotide or polypeptide, activity (or modulation of the activity) of the polynucleotide or polypeptide. In some cases, an alteration in a plant phenotype can be detected following contact of a plant (or plant cell, or tissue or explant) with the putative modulator, e.g., by modulation of expression or activity of a polynucleotide or polypeptide of the invention. Modulation of expression or activity of a polynucleotide or polypeptide of the invention may also be caused by molecular elements in a signal transduction second messenger pathway and such modulation can affect similar elements in the same or another signal transduction second messenger pathway.

XIV. Subsequences

Also contemplated are uses of polynucleotides, also referred to herein as oligonucleotides, typically having at least 12 bases, preferably at least 15, more preferably at least 20, 30, or 50 bases, which hybridize under at least highly stringent (or ultra-high stringent or ultra-ultra-high stringent conditions) conditions to a polynucleotide sequence described above. The polynucleotides may be used as probes, primers, sense and antisense agents, and the like, according to methods as noted *supra*.

Subsequences of the polynucleotides of the invention, including polynucleotide fragments and oligonucleotides are useful as nucleic acid probes and primers. An oligonucleotide suitable for use as a probe or primer is at least about 15 nucleotides in length, more often at least about 18 nucleotides, often at least about 21 nucleotides, frequently at least about 30 nucleotides, or about 40 nucleotides, or more in length. A nucleic acid probe is useful in hybridization protocols, e.g., to identify additional polypeptide homologues of the invention, including protocols for microarray experiments. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, e.g., by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods. See Sambrook and Ausubel, *supra*.

In addition, the invention includes an isolated or recombinant polypeptide including a subsequence of at least about 15 contiguous amino acids encoded by the recombinant or isolated polynucleotides of the invention. For example, such polypeptides, or domains or fragments thereof, can be used as immunogens, e.g., to produce antibodies specific for the polypeptide sequence, or as probes for detecting a sequence of interest. A subsequence can range in size from about 15 amino acids in length up to and including the full length of the polypeptide.

To be encompassed by the present invention, an expressed polypeptide which comprises such a polypeptide subsequence performs at least one biological function of the intact polypeptide in substantially the same manner, or to a similar extent, as does the intact polypeptide. For example, a polypeptide fragment can comprise a recognizable structural motif or functional domain such as a DNA binding domain that binds to a specific DNA promoter region, an activation domain or a domain for protein-protein interactions.

XV. Production of Transgenic Plants

Modification of Traits

The polynucleotides of the invention are favorably employed to produce transgenic plants with various traits, or characteristics, that have been modified in a desirable manner, e.g., to improve the seed characteristics of a plant. For example, alteration of expression levels or patterns (e.g., spatial or temporal expression patterns) of one or more of the transcription factors (or transcription factor homologues) of the invention, as compared with the levels of the same protein found in a wild type plant, can be used to modify a plant's traits. An illustrative example of trait modification, improved characteristics, by altering expression levels of a particular transcription factor is described further in the Examples and the Sequence Listing.

Arabidopsis as a model system

Arabidopsis thaliana is the object of rapidly growing attention as a model for genetics and metabolism in plants. *Arabidopsis* has a small genome, and well documented studies are available. It is easy to grow in large numbers and mutants defining important genetically controlled mechanisms are either available, or can readily be obtained. Various methods to introduce and express isolated homologous genes are available (see Koncz, et al., eds. *Methods in Arabidopsis Research*. et al. (1992), World Scientific, New Jersey, New Jersey, in "Preface"). Because of its small size, short life cycle, obligate autogamy and high fertility, *Arabidopsis* is also a choice organism for the isolation of mutants and studies in morphogenetic and development pathways, and control of these pathways by transcription factors (Koncz, *supra*, p. 72). A number of studies introducing transcription factors into *A. thaliana* have demonstrated the utility of this plant for understanding the mechanisms of gene regulation and trait alteration in plants. See, for example, Koncz, *supra*, and U.S. Patent Number 6,417,428).

Arabidopsis genes in transgenic plants.

Expression of genes which encode transcription factors modify expression of endogenous genes, polynucleotides, and proteins are well known in the art. In addition, transgenic plants comprising isolated polynucleotides encoding transcription factors may also modify expression of endogenous genes, polynucleotides, and

proteins. Examples include Peng et al. (1997, Genes and Development 11:3194-3205) and Peng et al. (1999, Nature, 400:256-261). In addition, many others have demonstrated that an *Arabidopsis* transcription factor expressed in an exogenous plant species elicits the same or very similar phenotypic response. See, for example, Fu et al. (2001, Plant Cell 13:1791-1802); Nandi et al. (2000, Curr. Biol. 10:215-218); Coupland (1995, Nature 377:482-483); and Weigel and Nilsson (1995, Nature 377:482-500).

Homologous genes introduced into transgenic plants.

Homologous genes that may be derived from any plant, or from any source whether natural, synthetic, semi-synthetic or recombinant, and that share significant sequence identity or similarity to those provided by the present invention, may be introduced into plants, for example, crop plants, to confer desirable or improved traits. Consequently, transgenic plants may be produced that comprise a recombinant expression vector or cassette with a promoter operably linked to one or more sequences homologous to presently disclosed sequences. The promoter may be, for example, a plant or viral promoter.

The invention thus provides for methods for preparing transgenic plants, and for modifying plant traits. These methods include introducing into a plant a recombinant expression vector or cassette comprising a functional promoter operably linked to one or more sequences homologous to presently disclosed sequences. Plants and kits for producing these plants that result from the application of these methods are also encompassed by the present invention.

The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4, Table 5, and Table 6.

Table 4

Polynucleotide SEQ ID NO:	GID No.	Trait	Category	Family	Comment	Polypeptide SEQ ID NO:	Conserved domains
1	G1275	Architecture; size	Dev and morph	WRKY	Reduced apical dominance; small plant	2	(113-169)
3	G1411	Architecture; light response; size; seed protein content	Dev and morph biochemistry	AP2	Loss of apical dominance	4	(87-154)
5	G1488	Architecture; flower; morphology; other	Dev and morph biochemistry	GATA/Zn	Reduced apical dominance, shorter stems; constitutive photomorphogenesis; reduced size; altered seed protein content	6	(221-246)
7	G1499	Architecture; flower; morphology; other; seed oil	Dev and morph biochemistry	HLH/MYC	Altered plant architecture; altered floral organ identity and development; dark green color	8	(118-181)
9	G1543	Architecture; morphology; other; fertility	Dev and morph biochemistry	HB	Altered plant architecture; altered carpel shape; dark green color; decreased seed oil	10	(135-195)
11	G1635	Architecture; light response; seed oil and protein content	Dev and morph biochemistry	MYB-related	Reduced apical dominance; pale green, smaller plants; reduced fertility	12	(44-104)
13	G1794	Architecture; size	Dev and morph biochemistry	AP2	Altered plant architecture; constitutive photomorphogenesis; altered seed oil and protein content	14	(182-248)
15	G1839	Architecture	Dev and morph	AP2	Altered plant architecture; reduced size	16	(118-184)
17	G2108	Architecture	Dev and morph	AP2	Altered inflorescence structure	18	(18-85)
19	G2291	Architecture; flowering time	Dev and morph flowering time	AP2	Altered plant architecture; late flowering	20	(TBD)
21	G2452	Architecture; leaf	Dev and morph	MYB-related	Reduced apical dominance; pale green color	22	(27-213)
23	G2509	Architecture; seed oil and protein content	Dev and morph biochemistry	AP2	Reduced apical dominance; altered seed oil and protein content	24	(89-156)
25	G390	Architecture	Dev and morph	HB	Altered shoot development	26	(18-81)
27	G391	Architecture	Dev and morph	HB	Altered shoot development	28	(25-85)
29	G438	Architecture; stem	Dev and morph	HB	Reduced branching; reduced lignin	30	(22-85)

Table 4

31	G47	Architecture; stem; flowering time; altered seed oil content	Dev and morph; flowering time; seed biochemistry	AP2	Altered architecture and inflorescence development, structure of vascular tissues; late flowering; altered seed oil content	32	(11-80)
33	G559	Architecture; fertility	Dev and morph; bZIP	bZIP	Loss of apical dominance; reduced fertility	34	(203-264)
35	G568	Architecture; flowering time	Dev and morph; bZIP	bZIP	Altered branching; late flowering	36	(215-265)
37	G580	Architecture; flower	Dev and morph; bZIP	bZIP	Altered inflorescences; altered flower development	38	(162-218)
39	G615	Architecture; fertility	Dev and morph; TEO	TEO	Altered plant architecture; little or no pollen production, poor filament elongation	40	(88-147)
41	G732	Architecture; flower; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Reduced apical dominance; abnormal flowers; altered seed oil and protein content	42	(31-91)
43	G988	Architecture; fertility; flower; stem; seed oil and protein content	Dev and morph; seed biochemistry	SCR	Reduced lateral branching; reduced fertility; enlarged floral organs, short pedicels; thicker stem, altered distribution of vascular bundles; altered seed oil and protein content	44	(178-195)
45	G1519	Embryo lethal	Dev and morph; RING/C3HC4	Embryo lethal		46	(327-364)
47	G374	Embryo lethal	Dev and morph; Z-ZPF	Embryo lethal		48	(35-67, 245-277)
49	G877	Embryo lethal	Dev and morph; WRKY	Embryo lethal		50	(272-328, 487-603)
51	G1000	Fertility; size; flower; stem	Dev and morph; MYB-(R1)R2R3	MYB-(R1)R2R3	Reduced fertility; small plant; reduced or absent petals and sepals; reduced inflorescence, stem elongation	52	(14-117)
53	G1067	Fertility; leaf; size	Dev and morph; AT-hook	AT-hook	Reduced fertility; altered leaf shape; small plant	54	(86-93)
55	G1075	Fertility; flower; leaf; size	Dev and morph; AT-hook	AT-hook	Reduced fertility; reduced or absent petals, sepals and stamens; altered leaf shape; small plant	56	(78-85)
57	G1266	Fertility; size	Dev and morph; AP2	AP2	Reduced fertility; small plant	58	(79-147)
59	G1311	Fertility; size	Dev and morph; MYB-(R1)R2R3	MYB-(R1)R2R3	Reduced fertility; small plant	60	(11-112)
61	G1321	Fertility; flower	Dev and morph; MYB-(R1)R2R3	MYB-(R1)R2R3	Poor fertility; altered flower morphology	62	(4-106)
63	G1326	Fertility; flower; size	Dev and morph; MYB-(R1)R2R3	MYB-(R1)R2R3	Reduced fertility; petals and sepals are smaller; small plant	64	(18-121)
65	G1367	Fertility; size	Dev and morph; AT-hook	AT-hook	Reduced fertility; reduced size	66	(179-201, 262-285, 298-319, 335-357)
67	G1386	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; reduced size; altered seed oil and protein content	68	(TBD)

Table 4

69	G1421	Fertility; size; seed oil content	Dev and morph; seed biochemistry	AP2	Reduced fertility; small plant; altered seed oil content	70	(74-151)
71	G1453	Fertility; morphology; other	Dev and morph	NAC	Reduced fertility; altered inflorescence development	72	(13-160)
73	G1560	Fertility; flower; size	Dev and morph	HS	Reduced fertility; altered flower development; reduced size	74	(62-151)
75	G1594	Fertility; leaf; seed	Dev and morph	HB	Reduced fertility; altered leaf shape and development; large pale seed	76	(343-308)
77	G1750	Fertility; size; seed oil content	Dev and morph; seed biochemistry	AP2	Reduced fertility; reduced size; increased seed oil content	78	(107-173)
79	G1947	Fertility; flower; seed protein content	Dev and morph; seed biochemistry	HS	Reduced fertility; extended period of flowering; altered seed protein content	80	(37-120)
81	G2011	Fertility; size; seed oil and protein content	Dev and morph; seed biochemistry	HS	Reduced fertility; reduced size; altered seed oil and protein content	82	(56-147)
83	G2094	Fertility; leaf; size	Dev and morph	GATA/Zn	Reduced fertility; altered leaf development; reduced size	84	(43-88)
85	G2113	Fertility; leaf; seed protein content	Dev and morph; seed biochemistry	AP2	Reduced fertility; long petioles, altered orientation; altered seed protein content	86	(TBD)
87	G2115	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	88	(46-115)
89	G2130	Fertility; size; senescence	Dev and morph	AP2	Reduced fertility; reduced size; early senescence	90	(93-160)
91	G2147	Fertility; size	Dev and morph	HLH/MYC	Reduced fertility; reduced size	92	(160-234)
93	G2156	Fertility; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced fertility; reduced size; altered seed protein content	94	(66-86)
95	G2294	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	96	(32-102)
97	G2510	Fertility; size	Dev and morph	AP2	Reduced fertility; reduced size	98	(41-108)
99	G2893	Fertility; flower; size	Dev and morph	MYB-(R1)/R2R3	Reduced fertility; altered flower development; reduced size	100	(19-120)
101	G340	Fertility; size	Dev and morph	Z-C3H	Reduced fertility, size	102	(37-154)
103	G39	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	104	(24-90)
105	G439	Fertility; size	Dev and morph	AP2	Reduced fertility, small plant	106	(110-177)
107	G470	Fertility	Dev and morph	ARF	Short stamen filaments	108	(61-393)

Table 4

109	G652	Fertility; seed; flower; size; seed oil content	Dev and morph; seed biochemistry	Z-CLDSH	Reduced fertility; irregular shaped seed; altered flower development; reduced size, slow growth; altered seed oil content	110	(28-49, 137-151, 182-196)
111	G671	Fertility; flower; leaf; size; stem	Dev and morph	MYB-(R1)R2R3	Reduced fertility; reduced petal abscission; altered leaf shape; small plant; altered inflorescence stem structure	112	(15-115)
113	G779	Fertility; flower	Dev and morph	HLH/MYC	Reduced fertility, homeotic transformations	114	(126-182)
115	G962	Fertility; size	Dev and morph	NAC	Reduced fertility; small plant	116	(53-175)
117	G977	Fertility; leaf; morphology; other; size	Dev and morph	AP2	Reduced fertility; altered leaf shape; dark green; small plant	118	(5-72)
119	G1063	Flower; leaf; inflorescence; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Altered flower development; ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development; altered seed oil and protein content	120	(131-182)
121	G1140	Flower	Dev and morph	MADS	Altered flower development	122	(2-57)
123	G1425	Flower	Dev and morph	NAC	Altered flower and inflorescence development	124	(20-173)
125	G1449	Flower	Dev and morph	IAA	Altered flower structure	126	(48-53, 74-107, 122-152)
127	G1897	Flower; leaf; seed protein content	Dev and morph; seed biochemistry	Z-Dof	Altered flower development; altered leaf development; altered seed protein content	128	(34-62)
129	G2143	Flower; leaf; inflorescence	Dev and morph	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color; altered inflorescence development	130	(128-179)
131	G2535	Flower; seed protein content	Dev and morph; seed biochemistry	NAC	Altered flower development; altered seed protein content	132	(11-114)
133	G2557	Flower; leaf	Dev and morph	HLH/MYC	Altered flower development, ectopic carpel tissue; altered leaf shape, dark green color	134	(278-328)
135	G259	Flower; leaf; size; seed protein content	Dev and morph	HS	Altered flower development; altered leaf development	136	(27-131)
137	G353	Flower; light response; size	Dev and morph; seed biochemistry	Z-C2H2	Short pedicels, downward pointing siliques; altered leaf development; reduced size; altered seed protein content	138	(41-61, 84-104)
139	G354	Flower; morphology; other	Dev and morph	Z-C2H2	Short pedicels, downward pointing siliques; constitutive morphogenesis; reduced size	140	(42-62, 88-109)
141	G538		Dev and morph	TH	Altered flower development; multiple developmental defects	142	(119-206)

Table 4

143	G869	Flower; morphology; other; seed oil	Dev and morph; biochemistry	AP2	Abnormal anther development; altered seed fatty acids	144	(109-177)
145	G1645	Inflorescence; leaf	Dev and morph	MYB-(R1)R2R3	Altered inflorescence structure; altered leaf development	146	(90-210)
147	G1038	Leaf	Dev and morph	GARP	Altered leaf shape	148	(198-247)
149	G1073	Leaf; size; flowering time	Dev and morph; flowering time	AT-hook	Serrated leaves; increased plant size; flowering appears to be slightly delayed	150	(33-42, 78-175)
151	G1146	Leaf	Dev and morph	PAZ	Altered leaf development	152	(886-896)
153	G1267	Leaf; size	Dev and morph	WRKY	Dark green shiny leaves; small plant	154	(70-127)
155	G1269	Leaf	Dev and morph	MYB-related	Long petioles, upturned leaves	156	(27-83)
157	G1452	Leaf; trichome; flowering time	Dev and morph; flowering time	NAC	Altered leaf shape, dark green color; reduced trichome density; late flowering	158	(30-177)
159	G1494	Leaf; size; light response; seed	Dev and morph	HLH/MYC	Pale green leaves, altered leaf shape; reduced size; long hypocotyls; large, pale seeds	160	(261-311)
161	G1548	Leaf	Dev and morph	HB	Altered leaf development	162	(17-77)
163	G1574	Leaf	Dev and morph	SWI/SNF	Altered leaf development	164	(28-350)
165	G1586	Leaf; size	Dev and morph	HB	Narrow leaves; small plants	166	(21-81)
167	G1786	Leaf; light response; size	Dev and morph	MYB-(R1)R2R3	Dark green, small leaves with short petioles; photomorphogenesis in the dark; small plant	168	(TBD)
169	G1792	Leaf; seed oil and protein content	Dev and morph; biochemistry	AP2	Dark green, shiny leaves; altered seed oil and protein content	170	(17-85)
171	G1865	Leaf; seed oil and protein content	Dev and morph; biochemistry	GRF-like	Altered leaf development; altered seed oil and protein content	172	(124-149)
173	G1886	Leaf; size	Dev and morph	Z-Dof	Chlorotic patches in leaves; reduced size	174	(17-59)
175	G1933	Leaf; size; seed protein content	Dev and morph; biochemistry	WRKY	Altered leaf development; reduced size; altered seed protein content	176	(205-263, 344-404)
177	G2059	Leaf; seed oil and protein content	Dev and morph; biochemistry	AP2	Smaller, curled leaves; altered seed oil, protein content	178	(184-254)
179	G2105	Leaf; seed	Dev and morph	TH	Alterations in leaf surface; large, pale seeds	180	(100-153)
181	G2117	Leaf; seed oil and protein content	Dev and morph; biochemistry	bZIP	Small, dark green leaves; altered seed oil and protein content	182	(46-106)

Table 4

183	G2124	Leaf; seed protein content	Dev and morph; seed biochemistry	TEO	Altered leaf development; altered seed protein content	184	(75-132)
185	G2140	Leaf; root	Dev and morph	HLH/MYC	Altered leaf development; short roots	186	(167-242)
187	G2144	Leaf; light response; size; seed oil content	Dev and morph; seed biochemistry	HLH/MYC	Pale green leaves, altered leaf shape; long hypocotyls; reduced size; altered seed oil content	188	(203-283)
189	G2431	Leaf	Dev and morph	GARP	Dark green leaves; reduced size	190	(38-88)
191	G2465	Morphology: other; leaf	Dev and morph	GARP	Slowed development; altered leaf color and shape	192	(219-269)
193	G2583	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Glossy, shiny leaves; altered seed oil and protein content	194	(4-71)
195	G2724	Leaf	Dev and morph	MYB-(R1)R2R3	Dark green leaves	196	(7-113)
197	G377	Morphology: other	Dev and morph	RING/C3H2C3	Altered leaf development; slow growth	198	(85-128)
199	G428	Leaf	Dev and morph	HB	Altered leaf shape	200	(229-292)
201	G447	Leaf; morphology: other; size	Dev and morph	ARF	Dark green leaves; altered cotyledon shape; reduced size	202	(22-356)
203	G464	Leaf	Dev and morph	IAA	Altered leaf shape	204	(20-28, 71-82, 126-142, 187-224)
205	G557	Leaf; size	Dev and morph	bZIP	Dark green color; small plant	206	(90-150)
207	G577	Leaf	Dev and morph	BZIP2	Reduced size; increased anthocyanins	208	(TBD)
209	G674	Leaf; size	Dev and morph	MYB-(R1)R2R3	Dark green leaves, upwardly oriented; reduced size	210	(20-120)
211	G736	Leaf; flowering time	Dev and morph; flowering time	Z-Dof	Altered leaf shape; later flowering	212	(54-111)
213	G903	Leaf	Dev and morph	Z-C2H2	Altered leaf morphology	214	(68-92)
215	G917	Leaf; seed oil and protein content	Dev and morph; seed biochemistry	MADS	Altered leaf development; altered seed oil and protein content	216	(2-57)
217	G921	Leaf	Dev and morph	WRKY	Serrated leaves	218	(146-203)
219	G922	Leaf; size	Dev and morph	SCR	Altered development, dark green color; reduced size	220	(225-242)
221	G932	Leaf; size	Dev and morph	MYB-(R1)R2R3	Altered development, dark green color; reduced size	222	(12-118)
223	G599	Leaf; size	Dev and morph	DBP	Altered leaf shape; small plant	224	(187-219, 264-300)
225	G804	Leaf; size	Dev and morph	PCF	Altered leaf shape, small plant	226	(54-117)

Table 4

227	G1062	Light response; morphology; other; seed	Dev and morph	HLH/MYC	Constitutive photomorphogenesis; slow growth; altered seed shape	228	(308-359)
229	G1322	Light response; size	Dev and morph	MYB-(R1)R2R3	Photomorphogenesis in the dark; reduced size	230	(26-130)
231	G1331	Light response; morphology; other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Constitutive photomorphogenesis; multiple developmental alterations; altered seed oil and protein content	232	(8-109)
233	G1521	Light response	Dev and morph	RING/C3HC4	Constitutive photomorphogenesis	234	(39-80)
235	G183	Light response; seed protein content	Dev and morph; seed biochemistry	WRKY	Constitutive photomorphogenesis; altered seed protein content	236	(307-363)
237	G2555	Light response	Dev and morph	HLH/MYC	Constitutive photomorphogenesis	238	(175-245)
239	G375	Light response	Dev and morph	Z-Dof	Upward pointing leaves	240	(75-103)
241	G1007	Morphology; other	Dev and morph	AP2	Multiple developmental alterations	242	(TBD)
243	G1010	Morphology; other	Dev and morph	ABI3/VP-1	Multiple developmental alterations	244	(33-122)
245	G1014	Morphology; other; trichome	Dev and morph	ABI3/VP-1	Multiple developmental defects; reduced trichomes	246	(90-172)
247	G1035	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	248	(39-91)
249	G1046	Morphology; other	Dev and morph	bZIP	Multiple developmental alterations	250	(79-138)
251	G1049	Morphology; other; seed protein content	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; altered seed protein content	252	(77-132)
253	G1069	Morphology; other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	254	(67-74)
255	G1070	Morphology; other	Dev and morph	AT-hook	Several developmental defects	256	(98-120)
257	G1076	Morphology; other	Dev and morph	AT-hook	Lethal when overexpressed	258	(82-89)
259	G1089	Morphology; other	Dev and morph	BZIPT2	Developmental defects at seedling stage	260	(425-500)
261	G1093	Morphology; other	Dev and morph	RING/C3H2C3	Multiple morphological alterations	262	(105-148)

Table 4

263	G1127	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	264	(103-110, 155-162)
265	G1131	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed protein content	266	(173-220)
267	G1145	Morphology: other; seed oil and protein	Dev and morph; seed biochemistry	bZIP	Multiple developmental alterations; reduced seed size, altered seed shape; altered seed oil and protein content	268	(227-270)
269	G1229	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Several developmental defects; altered seed oil and protein content	270	(102-160)
271	G1246	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	272	(27-139)
273	G1255	Morphology: other; seed	Dev and morph	Z-CO-like	Reduced apical dominance; increased seed size	274	(18-56)
275	G1304	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	276	(13-118)
277	G1318	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	278	(20-123)
279	G1320	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	280	(5-108)
281	G1330	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	282	(28-134)
283	G1352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	284	(108-129, 167-188)
285	G1354	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	286	(TBD)
287	G1360	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	288	(18-174)
289	G1364	Morphology: other	Dev and morph	CAAT	Lethal when overexpressed	290	(29-120)
291	G1379	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	292	(18-85)
293	G1384	Morphology: other	Dev and morph	AP2	Abnormal inflorescence and flower development	294	(TBD)
295	G1399	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	296	(86-93)

Table 4

297	G1415	Morphology: other	Dev and morph morph; seed biochemistry	AP2	Multiple developmental alterations	298	(TBD)
299	G1417	Morphology: other; seed oil		WRKY	Reduced seedling germination and vigor; increase in 18:2, decrease in 18:3	300	(239-296)
301	G1442	Morphology: other	Dev and morph	GRF-like	Multiple developmental alterations	302	(172-223)
303	G1454	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed oil and protein content	304	(9-178)
305	G1459	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	306	(10-152)
307	G1460	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	308	(TBD)
309	G147	Morphology: other	Dev and morph	MADS	Multiple developmental defects	310	(2-57)
311	G1471	Morphology: other; seed oil	Dev and morph; seed biochemistry	Z-C2H2	Multiple developmental alterations; increased seed oil content	312	(49-70)
313	G1475	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	314	(51-73)
315	G1477	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	316	(29-48)
317	G1487	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	318	(251-276)
319	G1492	Morphology: other	Dev and morph	GARP	Multiple developmental alterations	320	(34-83)
321	G1531	Morphology: other; seed; seed protein content	Dev and morph; seed biochemistry	RING/C3HC4	Multiple developmental alterations; pale seed; altered seed protein content	322	(41-77)
323	G1540	Morphology: other	Dev and morph	HB	Reduced cell differentiation in meristem	324	(35-98)
325	G1544	Morphology: other	Dev and morph	HB	Multiple developmental alterations	326	(64-124)

Table 4

327	G156	Morphology: other; seed	Dev and morph	MADS	Multiple developmental defects; seed color alteration	328	(2-57)
329	G1584	Morphology: other	Dev and morph	HB	Multiple developmental alterations	330	(TBD)
331	G1587	Morphology: other	Dev and morph	HB	Multiple developmental alterations	332	(61-121)
333	G1588	Morphology: other	Dev and morph	HB	Multiple developmental alterations	334	(66-124)
335	G1589	Morphology: other; seed protein content	Dev and morph; seed biochemistry	HB	Multiple developmental alterations; altered seed protein content	336	(384-448)
337	G160	Morphology: other	Dev and morph	MADS	Multiple developmental defects	338	(7-62)
339	G1636	Morphology: other	Dev and morph	MYB-related	Pale green, smaller plants	340	(100-165)
341	G1642	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	342	(TBD)
343	G1747	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed protein content	344	(11-114)
345	G1749	Morphology: other	Dev and morph	AP2	Multiple developmental alterations; formation of necrotic lesions	346	(84-155)
347	G1751	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	348	(TBD)
349	G1752	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	350	(83-151)
351	G1763	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	352	(140-209)
353	G1766	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	354	(10-153)
355	G1767	Morphology: other; seed oil content	Dev and morph; seed biochemistry	SCR	Multiple developmental alterations; altered seed oil content	356	(255-272)
357	G1778	Morphology: other	Dev and morph	GATA/Zn	Lethal when overexpressed	358	(94-119)
359	G1789	Morphology: other; seed protein content	Dev and morph; seed biochemistry	MYB-related	Delayed development; altered seed protein content	360	(1-50)
361	G1790	Morphology: other	Dev and morph	MYB-(R1)R2R3	Lethal when overexpressed	362	(217-316)

Table 4

363	G1791	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	364	(TBD)
365	G1793	Morphology: other; seed oil	Dev and morph; biochemistry	AP2	Multiple developmental alterations; increased seed oil content	366	(179-255, 281-349)
367	G1795	Morphology: other; trichome	Dev and morph	AP2	Multiple developmental alterations; reduced trichomes	368	(12-80)
369	G1800	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	370	(TBD)
371	G1806	Morphology: other	Dev and morph	bZIP	Multiple developmental alterations	372	(165-225)
373	G1811	Morphology: other	Dev and morph	AB13/V/P-1	Multiple developmental alterations	374	(TBD)
375	G182	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	376	(217-276)
377	G1835	Morphology: other	Dev and morph	GATA/Zn	Small, spindly plant	378	(224-296)
379	G1836	Morphology: other	Dev and morph	CAAT	Pale green	380	(30-164)
381	G1838	Morphology: other; seed oil content	Dev and morph; biochemistry	AP2	Multiple developmental alterations; increased seed oil content	382	(229-305, 330-400)
383	G1843	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	384	(2-57)
385	G1853	Morphology: other	Dev and morph	AKR	Lethal when overexpressed	386	(entire protein)
387	G1855	Morphology: other	Dev and morph	AKR	Slow growth	388	(entire protein)
389	G187	Morphology: other	Dev and morph	WRKY	Variety of morphological alterations	390	(172-228)
391	G1881	Morphology: other	Dev and morph	Z-CO-like	Multiple developmental alterations	392	(5-28, 56-79)
393	G1882	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	394	(97-125)
395	G1883	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	396	(82-124)
397	G1884	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	398	(43-71)
399	G1891	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	400	(27-69)

Table 4

401	G1896	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	402	(43-85)
403	G1898	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	404	(31-59)
405	G1902	Morphology: other; seed oil content	Dev and morph; seed biochemistry	Z-Dof	Multiple developmental alterations; increased seed oil content	406	(31-59)
407	G1904	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	408	(53-95)
409	G1906	Morphology: other	Dev and morph	Z-Dof	Multiple developmental alterations	410	(19-47)
411	G1913	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	412	(27-55)
413	G1914	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	414	(195-216, 245-266)
415	G1925	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	416	(6-150)
417	G1929	Morphology: other	Dev and morph	Z-CO-like	Slow growth, delayed development	418	(31-53)
419	G1930	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	420	(59-124)
421	G195	Morphology: other	Dev and morph	WRKY	Multiple developmental defects	422	(183-239)
423	G1954	Morphology: other	Dev and morph	HLH/MYC	Lethal when overexpressed	424	(187-259)
425	G1958	Morphology: other; seed protein content	Dev and morph; seed biochemistry	GARP	Reduced size and root mass in plates.; altered seed protein content	426	(230-278)
427	G196	Morphology: other; seed protein content	Dev and morph; seed biochemistry	WRKY	Multiple developmental alterations; altered seed protein content	428	(223-283)
429	G1965	Morphology: other	Dev and morph	Z-Dof	Lethal when overexpressed	430	(27-55)
431	G1976	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	432	(219-323)
433	G2057	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	434	(TBD)
435	G2107	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	436	(TBD)

Table 4

437	G211	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	438	(24-137)
439	G2133	Morphology: other; flowering time; seed protein content	Dev and morph; flowering time	AP2	Multiple developmental alterations; late flowering; altered seed protein content	440	(11-83)
441	G2134	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	442	(TBD)
443	G2151	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil and protein content	444	(93-113, 124-144)
445	G2154	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	446	(97-119)
447	G2157	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	448	(82-102, 164-107)
449	G2181	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	450	(22-169)
451	G221	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	452	(21-125)
453	G2290	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	454	(147-205)
455	G2299	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	456	(48-115)
457	G2340	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Tissue necrosis; multiple developmental alterations; altered seed oil and protein content	458	(14-120)
459	G2346	Morphology: other	Dev and morph	SBP	Enlarged seedlings	460	(59-135)
461	G237	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	462	(11-113)
463	G2373	Morphology: other; seed protein content	Dev and morph; seed biochemistry	TH	Multiple developmental alterations; altered seed protein content	464	(290-350)
465	G2376	Morphology: other; seed oil protein	Dev and morph; seed biochemistry	TH	Seedling lethality; altered seed protein content	466	(79-178, 336-408)
467	G24	Morphology: other	Dev and morph	AP2	Reduced size and necrotic patches	468	(25-93)

Table 4

469	G2424	Morphology: other	Dev and morph	MYB-(R1)R2R3	Multiple developmental alterations	470	(107-219)
471	G2505	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	472	(10-159)
473	G2512	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	474	(79-139)
475	G2513	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	476	(TBD)
477	G2519	Morphology: other	Dev and morph	HLH/MYC	Multiple developmental alterations	478	(1-65)
479	G2520	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Multiple developmental alterations; altered seed oil and protein content	480	(135-206)
481	G2533	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Multiple developmental alterations; altered seed protein content	482	(11-186)
483	G2534	Morphology: other	Dev and morph	NAC	Lethal when overexpressed	484	(10-157)
485	G2573	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed oil and protein content	486	(31-98)
487	G2589	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	488	(2-57)
489	G2687	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	490	(51-120)
491	G27	Morphology: other	Dev and morph	AP2	Abnormal development, small	492	(37-104)
493	G2720	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; altered seed oil and protein content	494	(10-114) (172-192, 226-247, 256-276, 290-311, 245 366)
495	G2787	Morphology: other; seed oil content	Dev and morph; seed biochemistry	AT-hook	Multiple developmental alterations; altered seed oil content	496	
497	G2789	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	498	(53-73, 121-165)
499	G31	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	500	(TBD)

Table 4

501	G33	Morphology: other	Dev and morph	AP2	Multiple developmental defects	502	(50-117)
503	G342	Morphology: other; seed oil and protein content	Dev and morph; seed biochemistry	GATA/Zn	Multiple developmental alterations; altered seed oil and protein content	504	(155-190)
505	G352	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	506	(99-119, 166-186)
507	G357	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	508	(7-29)
509	G358	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	510	(124-135, 188-210)
511	G360	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	512	(42-62)
513	G362	Size; Morphology: other; trichome; flowering time; seed protein content	Dev and morph; flowering time; seed biochemistry	Z-C2H2	Reduced size; increased pigmentation in seed, embryos and other organs; ectopic trichome formation; increased trichome number, late flowering; altered protein content	514	(62-82)
515	G364	Morphology: other	Dev and morph	Z-C2H2	Developmental defect	516	(54-76)
517	G365	Morphology: other	Dev and morph	Z-C2H2	Multiple developmental alterations	518	(70-90)
519	G367	Morphology: other	Dev and morph	Z-C2H2	Lethal when overexpressed	520	(63-84)
521	G373	Morphology: other	Dev and morph	RING/C3HC4	Multiple developmental alterations	522	(129-168)
523	G396	Morphology: other; size	Dev and morph	HB	Altered leaf coloration and shape, reduced fertility; small plant	524	(159-220)
525	G431	Morphology: other	Dev and morph	HB	Developmental defect, sterile	526	(286-335)
527	G479	Morphology: other	Dev and morph	SBP	Multiple developmental alterations	528	(70-149)
529	G546	Morphology: other	Dev and morph	RING/C3H2C3	Slow growth and development; increased anthocyanin pigmentation	530	(114-155)
531	G551	Morphology: other	Dev and morph	HB	Multiple developmental alterations	532	(73-133)
533	G578	Morphology: other	Dev and morph	bZIP	Lethal when overexpressed	534	(36-96)

Table 4

535	G596	Morphology: other	Dev and morph	AT-hook	Multiple developmental alterations	536	(89-96)
537	G617	Morphology: other	Dev and morph	TEO	Multiple developmental alterations	538	(64-118)
539	G620	Morphology: other; seed protein content	Dev and morph; seed biochemistry	CAAT	Multiple developmental alterations; altered seed protein content	540	(20-118)
541	G625	Morphology: other	Dev and morph	AP2	Lethal when overexpressed	542	(52-119)
543	G658	Morphology: other	Dev and morph	MYB-(R1)R2R3	Developmental defect	544	(2-105)
545	G716	Morphology: other	Dev and morph	ARF	Multiple developmental defects	546	(24-355)
547	G725	Morphology: other	Dev and morph	GARP	Developmental defect	548	(39-87)
549	G727	Morphology: other	Dev and morph	GARP	Multiple morphological alterations	550	(226-269)
551	G740	Morphology: other	Dev and morph	Z-CLDSH	Slow growth	552	(24-42, 232-268)
553	G770	Morphology: other	Dev and morph	NAC	Multiple developmental alterations	554	(19-162)
555	G858	Morphology: other	Dev and morph	MADS	Multiple developmental alterations	556	(2-57)
557	G865	Morphology: other; seed protein content	Dev and morph; seed biochemistry	AP2	Altered morphology; increased seed protein	558	(36-103)
559	G872	Morphology: other	Dev and morph	AP2	Multiple developmental alterations	560	(18-85)
561	G904	Morphology: other	Dev and morph	RING/C3H2C3	Multiple developmental alterations	562	(117-158)
563	G910	Morphology: other; flowering time	Dev and morph; flowering time	Z-CO-like	Multiple developmental alterations; late flowering	564	(14-37, 77-103)
565	G912	Morphology: other; size; sugar sensing; flowering time	Dev and morph; sensing; flowering time	AP2	Dark green color; small plant; reduced cotyledon expansion in glucose; late flowering	566	(51-118)
567	G920	Morphology: other	Dev and morph	WRKY	Multiple developmental alterations	568	(152-211)

Table 4

569	G939	Morphology: other; size	Dev and morph Dev and morph; seed biochemistry	EIL	Pale seedlings on agar; reduced size	570	(97-106)
571	G963	Morphology: other; seed protein content	Dev and morph; seed biochemistry	NAC	Slowed growth rate; altered seed protein content	572	(TBD)
573	G979	Morphology: other; seed	Dev and morph; seed biochemistry	AP2	Several developmental defects; altered seed development, ripening and germination	574	(63-139, 165-233)
575	G987	Morphology: other	Dev and morph; seed biochemistry	SCR	Developmental defects	576	(428-432, 704-708)
577	G993	Morphology: other; seed protein content	Dev and morph; seed biochemistry	AP2	Multiple developmental alterations; altered seed protein content	578	(69-134)
579	G681	Morphology: other; leaf glucosinolates	Dev and morph; leaf biochemistry	MYB-(R1)R2R3	Multiple developmental alterations; overexpression results in an increase in M39480	580	(14-120)
581	G1482	Root	Dev and morph; seed biochemistry	Z-CO-like	Increased root growth	582	(5-63)
583	G225	Root; trichome	Dev and morph; seed biochemistry	MYB-related	Increased root hairs; glabrous, lack of trichomes	584	(39-76)
585	G226	Root; trichome; seed protein content	Dev and morph; seed biochemistry	MYB-related	Increased root hairs; glabrous, lack of trichomes;	586	(28-78)
587	G9	Root	Dev and morph; seed biochemistry	AP2	Increased seed protein	588	(62-127)
589	G1040	Seed	Dev and morph; seed biochemistry	GARP	Increased root mass	590	(109-158)
591	G2114	Seed	Dev and morph; seed biochemistry	AP2	Smaller and more rounded seeds	592	(221-297, 323-393)
593	G450	Seed; size; seed protein content	Dev and morph; seed biochemistry	IAA	Increased seed size; reduced plant size; altered seed protein content	594	(TBD)
595	G584	Seed	Dev and morph; seed biochemistry	HLH/MYC	Large seeds	596	(401-494)
597	G668	Seed	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced seed color	598	(13-113)
599	G1050	Senescence	Dev and morph; seed biochemistry	bZIP	Delayed senescence	600	(372-425)
601	G1463	Senescence	Dev and morph; seed biochemistry	NAC	Premature senescence	602	(9-156)
603	G1944	Senescence; size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Early senescence; reduced size; altered seed protein content	604	(87-100)
605	G2383	Senescence; seed protein content	Dev and morph; seed biochemistry	TEO	Early senescence; altered seed protein content	606	(89-149)
607	G571	Senescence; flowering time	Dev and morph; seed biochemistry	bZIP	Delayed senescence; late flowering	608	(160-220)

Table 4

609	G636	Senescence; size	Dev and morph	TH	Premature senescence; reduced size	610	(55-145, 405-498)
611	G878	Senescence; flowering time	Dev and morph; flowering time	WRKY	Delayed senescence; late flowering	612	(250-305, 415-475)
613	G1134	Silique	Dev and morph	HLH/MYC	Siliques with altered shape	614	(198-247)
615	G1008	Size	Dev and morph	AP2	Small plant	616	(96-163)
617	G1020	Size	Dev and morph	AP2	Very small T1 plants	618	(28-95)
619	G1023	Size	Dev and morph	AP2	Reduced size	620	(128-195)
621	G1053	Size	Dev and morph	bZIP	Small plant	622	(74-120)
623	G1137	Size	Dev and morph	HLH/MYC	Small T1 plants	624	(264-314)
625	G1181	Size	Dev and morph	HS	Small T1 plants	626	(24-114)
627	G1228	Size	Dev and morph	HLH/MYC	Reduced size	628	(179-233)
629	G1277	Size	Dev and morph	AP2	Small plant	630	(18-85)
631	G1309	Size	Dev and morph	MYB-(R1)R2R3	Small plant	632	(9-114)
633	G1314	Size; sugar sensing; seed protein content	Dev and morph; sugar sensing; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced seedling vigor on high glucose; altered seed protein content	634	(14-116)
635	G1317	Size	Dev and morph	MYB-(R1)R2R3	Reduced size	636	(13-118)
637	G1323	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Small T1 plants, dark green; decreased seed oil, increased seed protein	638	(15-116)
639	G1332	Size; trichome; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; reduced trichome density; altered seed oil; and protein content	640	(13-116)
641	G1334	Size	Dev and morph	CAAT	Small, dark green	642	(18-190)
643	G1381	Size	Dev and morph	AP2	Reduced size	644	(68-135)
645	G1382	Size	Dev and morph	WRKY	Small plant	646	(210-266, 385-437)
647	G1435	Size; flowering time	Dev and morph; flowering time	GARP	Increased plant size; late flowering	648	(146-194)
649	G1537	Size	Dev and morph	HB	Small T1 plants with altered development	650	(14-74)
651	G1545	Size	Dev and morph	HB	Reduced size	652	(54-117)
653	G1641	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-related	Small plant; altered seed oil and protein content	654	(139-200)
655	G165	Size; seed protein content	Dev and morph; seed biochemistry	MADS	Reduced size; altered seed protein content	656	(7-62)

Table 4

657	G1652	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	658	(143-215)
659	G1655	Size	Dev and morph	HLH/MYC	Small plant	660	(134-192)
661	G1671	Size	Dev and morph	NAC	Reduced size	662	(TBD)
663	G1756	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Reduced size; altered seed protein content	664	(TBD)
665	G1757	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Small plant; altered seed protein content	666	(158-218)
667	G1782	Size	Dev and morph	CAAT	Small, spindly plant	668	(166-238)
669	G184	Size	Dev and morph	WRKY	Small plant	670	(295-352)
671	G1845	Size	Dev and morph	AP2	Small plant	672	(140-207)
673	G1879	Size; seed oil and protein content	Dev and morph; seed biochemistry	HLH/MYC	Reduced size; altered seed oil and protein content	674	(107-176)
675	G1888	Size	Dev and morph	Z-CO-like	Reduced size, dark green leaves	676	(5-50)
677	G189	Size; seed protein content	Dev and morph; seed biochemistry	WRKY	Increased leaf size; altered seed protein content	678	(240-297)
679	G1939	Size	Dev and morph	PCF	Reduced size	680	(40-102)
681	G194	Size	Dev and morph	WRKY	Small plant	682	(174-230)
683	G1943	Size	Dev and morph	HLH/MYC	Reduced size	684	(335-406)
685	G21	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	686	(97-164)
687	G2132	Size; seed oil and protein content	Dev and morph; seed biochemistry	AP2	Reduced size; altered seed oil and protein content	688	(TBD)
689	G2145	Size	Dev and morph	HLH/MYC	Reduced size	690	(166-243)
691	G23	Size	Dev and morph	AP2	Small T1 plants	692	(61-117)
693	G2313	Size	Dev and morph	MYB-related	Reduced size	694	(TBD)
695	G2344	Size	Dev and morph	CAAT	Reduced size, slow growth	696	(TBD)
697	G2430	Size	Dev and morph	GARP	Increased leaf size, faster development	698	(425-478)
699	G2517	Size	Dev and morph	WRKY	Reduced size	700	(118-234)
701	G2521	Size	Dev and morph	HLH/MYC	Reduced size	702	(145-213)
703	G258	Size; seed oil and protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Reduced size; altered seed oil and protein content	704	(24-124)

Table 4

705	G280	Size; seed protein content	Dev and morph; seed biochemistry	AT-hook	Reduced size; altered seed protein content	706	(97-104, 130-137-155-162, 185-192)
707	G3	Size	Dev and morph	AP2	Small plant	708	(28-95)
709	G343	Size	Dev and morph	GATA/Zn	Small plant	710	(178-214)
711	G363	Size	Dev and morph	Z-C2H2	Small plant	712	(87-108)
713	G370	Size	Dev and morph	Z-C2H2	Reduced size, shiny leaves	714	(97-117)
715	G385	Size	Dev and morph	HB	Small plant, short inflorescence stems, dark green	716	(60-123)
717	G439	Size	Dev and morph	AP2	Small plant	718	(110-177)
719	G440	Size	Dev and morph	AP2	Small plant	720	(122-189)
721	G5	Size	Dev and morph	AP2	Small plant	722	(149-216)
723	G550	Size	Dev and morph	Z-Dof	Small plant	724	(134-180)
725	G670	Size	Dev and morph	MYB-(R1)R2R3	Small plant	726	(14-122)
727	G760	Size	Dev and morph	NAC	Reduced size	728	(12-156)
729	G831	Size	Dev and morph	AKR	Reduced size	730	(470-591)
731	G864	Size	Dev and morph	AP2	Small plant	732	(119-186)
733	G884	Size	Dev and morph	WRKY	Reduced size	734	(227-285, 407-465)
735	G898	Size; seed oil and protein content	Dev and morph; seed biochemistry	RING/C3HC4	Reduced size; altered seed oil and protein content	736	(148-185)
737	G900	Size	Dev and morph	Z-CO-like	Reduced size	738	(6-28, 48-74)
739	G913	Size; flowering time	Dev and morph; flowering time	AP2	Small plant; late flowering	740	(62-128)
741	G937	Size	Dev and morph	GARP	Slightly reduced size	742	(197-246)
743	G960	Size	Dev and morph	NAC	Small plant	744	(13-156)
745	G991	Size; seed oil and protein content	Dev and morph; seed biochemistry	IAA	Slightly reduced size; altered seed oil and protein content	746	(7-14, 48-59, 82-115, 128-164)
747	G748	Stem; flowering time	Dev and morph; flowering time	Z-Dof	More vascular bundles in stem; late flowering	748	(112-140)
749	G247	Trichome; seed protein content	Dev and morph; seed biochemistry	MYB-(R1)R2R3	Altered trichome distribution; altered seed protein content	750	(15-116)
751	G585	Trichome	Dev and morph	HLH/MYC	Reduced trichome density	752	(436-501)
753	G634	Trichome; seed protein content	Dev and morph; seed biochemistry	TH	Increased trichome density and size; altered seed protein content	754	(62-147, 189-245)
755	G676	Trichome	Dev and morph	MYB-(R1)R2R3	Reduced trichomes	756	(17-119)

Table 4

757	G682	Trichome	Dev and morph	MYB-related	Glabrous, lack of trichomes	758	(27-63)
759	G635	Variegation	Dev and morph	TH		760	(239-323)
761	G1068	Sugar sensing; seed oil and protein content	Sugar sensing; seed	AT-hook	Reduced cotyledon expansion in glucose	762	(143-150)
763	G1225	Sugar sensing; seed oil and protein content	Sugar sensing; seed	HLH/MYC	Better germination on sucrose and glucose media; altered seed oil and protein content	764	(78-147)
765	G1337	Sugar sensing	Sugar sensing	Z-CO-like	Decreased germination on sucrose medium	766	(9-75)
767	G1759	Sugar sensing	Sugar sensing	MADS	Reduced germination on high glucose	768	(2-57)
769	G1804	Sugar sensing; flowering time	Sugar sensing; flowering time	bZIP	Altered sugar sensing; late flowering	770	(357-407)
771	G207	Sugar sensing	Sugar sensing	MYB-(R1)R2R3	Decreased germination on glucose medium	772	(6-106)
773	G218	Sugar sensing; seed oil content	Sugar sensing; seed	MYB-(R1)R2R3	Reduced cotyledon expansion in glucose; altered seed oil content	774	(TBD)
775	G241	Sugar sensing; seed oil and protein content	Sugar sensing; seed	MYB-(R1)R2R3	Decreased germination and growth on glucose medium; decreased seed oil, altered protein content	776	(14-114)
777	G254	Sugar sensing	Sugar sensing	MYB-related	Decreased germination and growth on glucose medium	778	(62-106)
779	G26	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	780	(67-134)
781	G263	Sugar sensing	Sugar sensing	HS	Decreased root growth on sucrose medium, root specific expression	782	(TBD)
783	G308	Sugar sensing	Sugar sensing	SCR	No germination on glucose medium	784	(270-274)
785	G38	Sugar sensing	Sugar sensing	AP2	Reduced germination on glucose medium	786	(76-143)
787	G43	Sugar sensing	Sugar sensing	AP2	Decreased germination and growth on glucose medium	788	(104-172)
789	G536	Sugar sensing	Sugar sensing	GF14	Decreased germination and growth on glucose medium	790	(226-233)
791	G567	Sugar sensing; seed oil and protein content	Sugar sensing; seed	bZIP	Decreased seedling vigor on high glucose; altered seed oil and protein content	792	(210-270)
793	G680	Sugar sensing; flowering time	Sugar sensing; flowering time	MYB-related	Reduced germination on glucose medium; late flowering	794	(24-70)
795	G867	Sugar sensing	Sugar sensing	AP2	Better seedling vigor on sucrose medium	796	(59-124)
797	G956	Sugar sensing	Sugar sensing	NAC	Reduced germination on glucose medium	798	(TBD)
799	G996	Sugar sensing	Sugar sensing	MYB-(R1)R2R3	Reduced germination on glucose medium	800	(14-114)
801	G1946	Seed glucosinolates, oil, protein content	Seed biochemistry	HS	Increase in M3950; increased oil content; decreased protein content	802	(32-130)
803	G217	Seed oil composition	Seed biochemistry	MYB-related	Increase in 20:2	804	(8-67)

Table 4

805	G2192	Seed oil composition	Seed biochemistry	bZIP-NIN	Altered composition	806	(600-700)
807	G504	Seed oil composition;	Seed biochemistry	NAC	Altered seed oil composition and content; altered seed protein content	808	(TBD)
809	G622	Seed oil composition	Seed biochemistry	ABI3/NP-1	Decreased 18:2 fatty acid	810	(TBD)
811	G778	Seed oil composition	Seed biochemistry	HLH/MYC	Increased seed 18:1 fatty acid	812	(220-267)
813	G791	Seed oil composition	Seed biochemistry	HLH/MYC	Altered seed fatty acid composition	814	(75-143)
815	G861	Seed oil composition; seed oil content	Seed biochemistry	MADS	Increase in 16:1; altered seed oil content	816	(2-57)
817	G938	Seed oil composition	Seed biochemistry	EIL	Altered seed fatty acid composition	818	(96-104)
819	G965	Seed oil composition	Seed biochemistry	HB	Increase in 18:1	820	(423-486)
821	G1143	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	822	(33-82)
823	G1190	Seed oil content	Seed biochemistry	AKR	Increased content	824	(entire protein)
825	G1198	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	826	(173-223)
827	G1226	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	828	(115-174)
829	G1451	Seed oil content	Seed biochemistry	ARF	Altered seed oil content	830	(22-357)
831	G1478	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	Z-CO-like	Altered seed oil, protein content; late flowering	832	(32-76)
833	G1496	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	834	(184-248)
835	G1526	Seed oil content	Seed biochemistry	SWI/SNF	Increased seed oil content	836	(493-620, 864-1006)
837	G1543	Seed oil content	Seed biochemistry	HB	Decreased seed oil	838	(135-195)
839	G162	Seed oil and protein content	Seed biochemistry	MADS	Altered seed oil content; altered seed oil and protein content	840	(2-57)

Table 4

841	G1640	Seed oil content; Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Increased seed oil	842	(14-115)
843	G1644	Seed oil and protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil, protein content	844	(39-102)
845	G1646	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	846	(72-162)
847	G1672	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	848	(41-194)
849	G1677	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil, protein content	850	(17-181)
851	G1765	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	852	(20-140)
853	G1777	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Increased oil, decreased protein content	854	(124-247)
855	G1793	Seed oil content	Seed biochemistry	AP2	Increased seed oil content	856	(179-255, 281-349)
857	G180	Seed oil content	Seed biochemistry	WRKY	Decreased seed oil content	858	(118-174)
859	G192	Seed oil and protein content; flowering time	Seed biochemistry; flowering time	WRKY	Altered seed oil and protein content; late flowering	860	(128-185)
861	G1948	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	862	(entire protein)
863	G2123	Seed oil and protein content	Seed biochemistry	GF14	Altered seed oil and protein content	864	(99-109)
865	G2138	Seed oil content	Seed biochemistry	AP2	Increased seed oil content	866	(TBD)
867	G2139	Seed oil content	Seed biochemistry	MADS	Increased seed content	868	(14-69)
869	G2343	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	870	(14-116)
871	G265	Seed oil and protein content	Seed biochemistry	HS	Altered seed oil and protein content	872	(11-105)
873	G2792	Seed oil content	Seed biochemistry	HLH/MYC	Increased seed oil content	874	(190-258)
875	G2830	Seed oil and protein content	Seed biochemistry	Z-C2H2	Altered seed oil and protein content	876	(245-266)
877	G286	Seed oil and protein content	Seed biochemistry	ENBP	Altered seed oil and protein content	878	(TBD)

Table 4

879	G291	Seed oil content	Seed biochemistry	MISC	Increased seed oil content	880	(132-160)
881	G427	Seed oil and protein content	Seed biochemistry	HB	Increased oil content; decreased protein content	882	(307-370)
883	G509	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	884	(13-169)
885	G519	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	886	(11-104)
887	G561	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	888	(248-308)
889	G590	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	890	(202-254)
891	G818	Seed oil content	Seed biochemistry	HS	Increased content	892	(70-162)
893	G849	Seed oil and protein content	Seed biochemistry	BPF-1	Increased seed oil, altered protein content	894	(324-413, 504-583)
895	G892	Seed oil and protein content	Seed biochemistry	RING/C3H2C3	Altered seed oil, protein content	896	(177-270)
897	G961	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	898	(15-140)
899	G1465	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	900	(242-306)
901	G425	Seed oil content	Seed biochemistry	HB	Altered seed oil content	902	(TBD)
903	G347	Seed oil and protein content	Seed biochemistry	Z-LSDlike	Altered seed oil and protein content	904	(9-39, 50-70, 80-127)
905	G1512	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	906	(39-93)
907	G2069	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	908	(TBD)
909	G1852	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	910	(1-601)
911	G1793	Seed oil content	Seed biochemistry	AP2	Altered seed oil content	912	(179-255, 281-349)
913	G761	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	914	(10-156)
915	G1056	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	916	(183-246)
917	G1447	Seed oil content	Seed biochemistry	MISC	Altered seed oil content	918	(3-54, 124-156)

Table 4

919	G323	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	920	(48-96)
921	G176	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	922	(117-173, 234-290)
923	G174	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	924	(111-166, 283-339)
925	G715	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	926	(60-132)
927	G588	Seed oil and protein content	Seed biochemistry	HLH/MYC	Altered seed oil and protein content	928	(309-376)
929	G1758	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	930	(109-165)
931	G2148	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	932	(130-268)
933	G2379	Seed oil content	Seed biochemistry	TH	Altered seed oil content	934	(19-110, 173-232)
935	G1462	Seed oil content	Seed biochemistry	NAC	Altered seed oil content	936	(TBD)
937	G1211	Seed oil and protein content	Seed biochemistry	MISC	Altered seed oil and protein content	938	(123-179)
939	G1048	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	940	(138-190)
941	G986	Seed oil content	Seed biochemistry	WRKY	Altered seed oil content	942	(146-203)
943	G789	Seed oil content	Seed biochemistry	HLH/MYC	Altered seed oil content	944	(253-313)
945	G2085	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	946	(TBD)
947	G1783	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	948	(81-129)
949	G2072	Seed oil and protein content	Seed biochemistry	bZIP	Altered seed oil and protein content	950	(90-149)
951	G931	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil and protein content	952	(TBD)
953	G278	Seed oil and protein content	Seed biochemistry	AKR	Altered seed oil and protein content	954	(2-593)
955	G2421	Seed oil content	Seed biochemistry	MYB-(R1)R2R3	Altered seed oil content	956	(9-110)
957	G2032	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	958	(entire protein)

Table 4

959	G1396	Seed oil and protein content	Seed biochemistry	SIFA	Altered seed oil and protein content	960	(TBD)
961	G819	Seed oil and protein content	Seed biochemistry	ARF	Altered seed oil and protein content	962	(64-406)
963	G2295	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	964	(2-57)
965	G312	Seed oil content	Seed biochemistry	SCR	Altered seed oil content	966	(320-336)
967	G1444	Seed oil and protein content	Seed biochemistry	GRF-like	Altered seed oil and protein content	968	(168-193)
969	G801	Seed oil content	Seed biochemistry	PCF	Altered seed oil content	970	(32-93)
971	G1950	Seed oil content	Seed biochemistry	AKR	Altered seed oil content	972	(65-228)
973	G958	Seed oil and protein content	Seed biochemistry	NAC	Altered seed oil and protein content	974	(7-156)
975	G1037	Seed oil and protein content	Seed biochemistry	GARP	Altered seed oil and protein content	976	(11-134, 200-248)
977	G2065	Seed oil content	Seed biochemistry	MADS	Altered seed oil content	978	(TBD)
979	G2137	Seed oil and protein content	Seed biochemistry	WRKY	Altered seed oil and protein content	980	(109-168)
981	G746	Seed oil content	Seed biochemistry	RING/C3HC4	Altered seed oil content	982	(139-178)
983	G2701	Seed oil and protein content	Seed biochemistry	MYB-related	Altered seed oil and protein content	984	(33-81, 129-183)
985	G1819	Seed oil content	Seed biochemistry	CAAT	Altered seed oil content	986	(46-188)
987	G1227	Seed oil and protein content	Seed biochemistry	HLL/MYC	Altered seed oil and protein content	988	(183-244)
989	G2417	Seed oil content	Seed biochemistry	GARP	Altered seed oil content	990	(235-285)
991	G2116	Seed oil content	Seed biochemistry	bZIP	Altered seed oil content	992	(150-210)
993	G647	Seed oil content	Seed biochemistry	Z-C3H	Altered seed oil content	994	(77-192)
995	G974	Seed oil and protein content	Seed biochemistry	AP2	Altered seed oil and protein content	996	(81-140)
997	G1419	Seed protein content	Seed biochemistry	AP2	Increased seed protein	998	(69-137)

Table 4

999	G1634	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1000	(129-180)
1001	G1637	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1002	(109-173)
1003	G1818	Seed protein content; flowering time	Seed biochemistry; flowering time	CAAT	Increased protein content; late flowering	1004	(36-113)
1005	G1820	Seed oil and protein content	Seed biochemistry	CAAT	Altered seed oil, protein content	1006	(70-133)
1007	G1903	Seed oil and protein content	Seed biochemistry	Z-Dof	Altered seed oil and protein content	1008	(134-180)
1009	G371	Seed oil and protein content	Seed biochemistry	RING/C3HC4	Altered seed oil and protein content	1010	(21-74)
1011	G597	Seed protein content	Seed biochemistry	AT-hook	Altered seed protein content	1012	(97-104, 137-144)
1013	G1009	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1014	(201-277, 303-371)
1015	G170	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1016	(2-57)
1017	G1768	Seed protein content	Seed biochemistry	SCR	Altered seed protein content	1018	(54-413)
1019	G185	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1020	(113-172)
1021	G1931	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1022	(114-170)
1023	G2543	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1024	(31-91)
1025	G264	Seed protein content	Seed biochemistry	HS	Altered seed protein content	1026	(24-114)
1027	G32	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1028	(17-84)
1029	G436	Seed protein content	Seed biochemistry	HB	Altered seed protein content	1030	(22-85)
1031	G556	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1032	(83-143)
1033	G1420	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1034	(221-280)
1035	G1412	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1036	(17-159)

Table 4

1037	G738	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1038	(351-393)
1039	G2426	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1040	(14-114)
1041	G1524	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1042	(49-110)
1043	G1243	Seed protein content	Seed biochemistry	SWI/SNF	Altered seed protein content	1044	(216-609)
1045	G631	Seed protein content	Seed biochemistry	bZIP	Altered seed protein content	1046	(TBD)
1047	G1909	Seed protein content	Seed biochemistry	Z-Dof	Altered seed protein content	1048	(23-51)
1049	G1663	Seed protein content	Seed biochemistry	PCF	Altered seed protein content	1050	(TBD)
1051	G1231	Seed protein content	Seed biochemistry	Z-C4HC3	Altered seed protein content	1052	(TBD)
1053	G227	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1054	(13-112)
1055	G1842	Seed protein content	Seed biochemistry	MADS	Altered seed protein content	1056	(2-57)
1057	G1505	Seed protein content	Seed biochemistry	GATA/Zn	Altered seed protein content	1058	(TBD)
1059	G657	Seed protein content	Seed biochemistry	MYB-(R1)R2R3	Altered seed protein content	1060	(TBD)
1061	G1959	Seed protein content	Seed biochemistry	GARP	Altered seed protein content	1062	(46-97)
1063	G2180	Seed protein content	Seed biochemistry	NAC	Altered seed protein content	1064	(7-156)
1065	G1817	Seed protein content	Seed biochemistry	PMR	Altered seed protein content	1066	(47-331)
1067	G1649	Seed protein content	Seed biochemistry	HLH/MYC	Altered seed protein content	1068	(225-295)
1069	G2131	Seed protein content	Seed biochemistry	AP2	Altered seed protein content	1070	(50-186, 112-183)
1071	G215	Seed protein content	Seed biochemistry	MYB-related	Altered seed protein content	1072	(TBD)
1073	G1508	Seed protein content	Seed biochemistry	GATA/Zn	Altered seed protein content	1074	(38-63)
1075	G2110	Seed protein content	Seed biochemistry	WRKY	Altered seed protein content	1076	(239-298)

Table 4

1077	G2442	Seed protein content	Seed biochemistry	RING/C3HC4	Altered seed protein content	1078	(220-246)
1079	G1051	Flowering time	Flowering time	bZIP	Late flowering	1080	(189-250)
1081	G1052	Flowering time	Flowering time	bZIP	Late flowering	1082	(201-261)
1083	G1079	Flowering time	Flowering time	BZIP12	Late flowering; altered seed protein content	1084	(1-50)
1085	G1335	Flowering time	Flowering time	Z-CLDSH	Late flowering, slow growth	1086	(24-43, 131-144, 185-203)
1087	G157	Flowering time	Flowering time	MADS	Altered flowering; significant overexpression delays	1088	(2-57)
1089	G1895	Flowering time	Flowering time	Z-Dof	Late flowering	1090	(55-110)
1091	G1900	Flowering time	Flowering time	Z-Dof	Late flowering	1092	(54-106)
1093	G2007	seed protein content	seed biochemistry	MYB-(R1)R2R3	Late flowering; altered seed protein content	1094	(TBD)
1095	G214	Flowering time	Flowering time	MYB-related	Late flowering	1096	(22-71)
1097	G2155	Flowering time	Flowering time	AT-hook	Late flowering	1098	(18-38)
1099	G234	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering, small plant	1100	(14-115)
1101	G361	Flowering time	Flowering time	Z-C2H2	Late flowering	1102	(43-63)
1103	G562	Flowering time	Flowering time	bZIP	Late flowering	1104	(253-315)
1105	G591	Flowering time	Flowering time	HLH/MYC	Late flowering	1106	(143-240)
1107	G8	Flowering time	Flowering time	AP2	Late flowering	1108	(151-217, 243-296)
1109	G859	seed protein content	seed biochemistry	MADS	Late flowering; altered seed protein content	1110	(TBD)
1111	G878	Flowering time	Flowering time	WRKY	Late flowering	1112	(250-305, 415-475)
1113	G971	Flowering time	Flowering time	AP2	Late flowering	1114	(120-186)
1115	G975	Flowering time; morphology: other	Flowering time; dev and morph	AP2	Late flowering; glossy leaves	1116	(4-71)
1117	G994	Flowering time	Flowering time	MYB-(R1)R2R3	Late flowering, small	1118	(14-123)
1119	G2347	Flowering time	Flowering time	SBP	Late flowering, small	1120	(60-136)
1121	G2010	Flowering time	Flowering time	SBP	Late flowering	1122	(53-127)

Table 5

SEQ ID NO	GID	Test Sequence ID	Smallest Sum Probability	Test Sequence Species	Test Sequence GenBank Annotation
859	G192	AW596933	7.70E-40	[Glycine max]	s184f07.y1 Gm-c1034 Glycine max cDNA clone GENO
859	G192	AV423663	2.40E-39	[Lotus japonicus]	AV423663 Lotus japonicus young plants (two-
859	G192	BI422074	4.50E-34	[Lycopersicon esculentum]	EST532740 tomato callus, TAMU Lycop
859	G192	AW447931	1.40E-27	[Triticum aestivum]	BRY_1082 BRY Triticum aestivum cDNA clone
859	G192	BE998060	2.60E-24	[Medicago truncatula]	EST429783 GVSJN Medicago truncatula cDNA
859	G192	AC018727	1.70E-23	[Oryza sativa]	chromosome 10 clone OSJNBa0056G17, *** SEQUENC
859	G192	BG600477	1.00E-20	[Solanum tuberosum]	EST505372 cSTS Solanum tuberosum cDNA clo
859	G192	BG356878	2.80E-16	[Sorghum bicolor]	OV2_11_B04.g1_A002 Ovary 2 (OV2) Sorghum bi
859	G192	gi12039364	1.10E-31	[Oryza sativa]	putative DNA-binding protein.
859	G192	gi4894963	3.30E-14	[Avena sativa]	DNA-binding protein WRKY3.
859	G192	gi1432056	5.80E-14	[Petroselinum crispum]	WRKY3.
859	G192	gi4760596	2.60E-13	[Nicotiana tabacum]	DNA-binding protein NWRKY3.
859	G192	gi11993901	1.40E-12	[Dactylis glomerata]	somatic embryogenesis related protein.
859	G192	gi927025	7.60E-09	[Cucumis sativus]	SPF1-like DNA-binding protein.
859	G192	gi13620227	8.40E-09	[Lycopersicon esculentum]	hypothetical protein.
859	G192	gi3420906	2.80E-08	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
859	G192	gi1159877	4.70E-08	[Avena fatua]	DNA-binding protein.
859	G192	gi484261	1.60E-07	[Ipomoea batatas]	SPF1 protein.
801	G1946	LPHSF8	1.10E-119	[Lycopersicon peruvianum]	L.peruvianum Lp-hsf8 mRNA for heat
801	G1946	AC087771	4.10E-112	[Medicago truncatula]	clone 8D15, *** SEQUENCING IN PROGRESS
801	G1946	LEHSF8	5.90E-103	[Lycopersicon esculentum]	L.esculentum Le-hsf8 gene for heat
801	G1946	AW569138	3.10E-75	[Glycine max]	s163g09.y1 Gm-r1030 Glycine max cDNA clone GENO
801	G1946	BG890899	1.30E-70	[Solanum tuberosum]	EST516750 cSTD Solanum tuberosum cDNA clo
801	G1946	AC027658	4.60E-53	[Oryza sativa]	subsp. japonica BAC nbx0006113, chromosome 10
801	G1946	AV833112	4.90E-52	[Hordeum vulgare subsp. vulgare]	AV833112 K. Sato unpublished
801	G1946	gi19492	2.80E-121	[Lycopersicon peruvianum]	heat shock transcription factor 8
801	G1946	gi19260	5.10E-106	[Lycopersicon esculentum]	heat stress transcription factor
801	G1946	gi662924	2.00E-47	[Glycine max]	heat shock transcription factor 21.
801	G1946	gi5821138	9.70E-46	[Nicotiana tabacum]	heat shock factor.
801	G1946	gi11761077	2.90E-40	[Oryza sativa]	putative heat shock factor protein 1 (HSF 1)
801	G1946	gi886742	3.20E-40	[Zea mays]	heat shock factor.
801	G1946	gi7158882	2.70E-38	[Medicago sativa]	heat shock transcription factor.
801	G1946	gi3550588	1.90E-30	[Pisum sativum]	heat shock transcription factor (HSFA).

Table 5

801	G1946	gi100546	0.46	[Avena sativa]	avenin precursor - oat.
801	G1946	gi14190783	1	[Apium graveolens]	putative phloem transcription factor M1.
239	G375	AW696439	3.40E-33	[Medicago truncatula]	NF106B07ST1F1060 Developing stem Medica
239	G375	BG595870	1.90E-31	[Solanum tuberosum]	EST494548 cSTS Solanum tuberosum cDNA clo
239	G375	AI899263	3.70E-31	[Lycopersicon esculentum]	EST288706 tomato ovary, TAMU Lycope
239	G375	NTBBF3	4.00E-31	[Nicotiana tabacum]	N.tabacum mRNA for zinc finger protein, B
239	G375	BG405482	2.70E-30	[Glycine max]	sac44a11.y1 Gm-c1062 Glycine max cDNA clone GEN
239	G375	AB028130	3.30E-30	[Oryza sativa]	mRNA for Dof zinc finger protein, complete cds
239	G375	AB026297	7.30E-28	[Pisum sativum]	mRNA for elicitor-responsive Dof protein ERDP
239	G375	HVBPBF	1.10E-27	[Hordeum vulgare]	mRNA for DNA binding protein BPBF.
239	G375	BG263089	1.70E-27	[Triticum aestivum]	WHE2337_A02_A03ZS Wheat pre-anthesis spik
239	G375	ZMU82230	4.20E-27	[Zea mays]	endosperm-specific prolamin box binding factor (PB
239	G375	gi4996640	1.90E-37	[Oryza sativa]	Dof zinc finger protein.
239	G375	gi3777436	8.10E-35	[Hordeum vulgare]	DNA binding protein.
239	G375	gi2393775	1.10E-33	[Zea mays]	prolamin box binding factor.
239	G375	gi1360088	2.00E-33	[Nicotiana tabacum]	Zn finger protein.
239	G375	gi3790264	4.30E-32	[Triticum aestivum]	PBF protein.
239	G375	gi6092016	1.30E-29	[Pisum sativum]	elicitor-responsive Dof protein ERDP.
239	G375	gi7688355	5.60E-29	[Solanum tuberosum]	Dof zinc finger protein.
239	G375	gi1669341	4.60E-20	[Cucurbita maxima]	AOBP (ascorbate oxidase promoter-binding
239	G375	gi3929325	5.50E-18	[Dendrobium grex Madame Thong-In]	putative DNA-binding prot
239	G375	gi19547	5.50E-06	[Medicago sativa subsp. falcata]	environmental stress and a
273	G1255	AC087181	1.60E-46	[Oryza sativa]	chromosome 3 clone OSJNBa0018H01, *** SEQUENCI
273	G1255	BG239774	4.50E-33	[Glycine max]	sab74c03.y1 Gm-c1032 Glycine max cDNA clone GEN
273	G1255	BG321336	1.70E-32	[Descurainia sophia]	Ds01_06h10_A Ds01_AAFc ECORC cold stress
273	G1255	AI772841	2.90E-30	[Lycopersicon esculentum]	EST253941 tomato resistant, Cornell
273	G1255	BF480245	4.60E-29	[Mesembryanthemum crystallinum]	L0-2152T3 Ice plant Lambda Un
273	G1255	AW688119	2.10E-28	[Medicago truncatula]	NF002E07ST1F1000 Developing stem Medica
273	G1255	BF266327	1.80E-26	[Hordeum vulgare]	HV_CEA0014N02f Hordeum vulgare seedling gre
273	G1255	AW671538	5.80E-25	[Sorghum bicolor]	LG1_348_B08.b1_A002 Light Grown 1 (LG1) Sor
273	G1255	BI072021	5.30E-20	[Populus tremula x Populus tremuloides]	C067P76U Populus stra
273	G1255	BG273908	4.90E-19	[Vitis vinifera]	EST 110 Green Grape berries Lambda Zap II Li
273	G1255	gi13702811	3.70E-52	[Oryza sativa]	putative zinc finger protein.
273	G1255	gi11037311	4.00E-21	[Brassica nigra]	constans-like protein.
273	G1255	gi2303683	1.10E-19	[Brassica napus]	unnamed protein product.
273	G1255	gi4091804	2.30E-18	[Malus x domestica]	CONSTANS-like protein 1.

Table 5

273	G1255	gi3341723	4.30E-18	[Raphanus sativus]	CONSTANS-like 1 protein.
273	G1255	gi10946337	5.20E-17	[Ipomoea nil]	CONSTANS-like protein.
273	G1255	gi4557093	3.30E-15	[Pinus radiata]	zinc finger protein.
273	G1255	gi8132543	0.97	[Chloroplast Zamia furfuracea]	cytochrome b559 alpha subunit
273	G1255	gi111795	0.99	[Nicotiana tabacum]	put. psbE protein (aa 1-83).
273	G1255	gi65646	0.99	[Chloroplast Nicotiana tabacum]	cytochrome b559 component p
557	G865	BE419451	3.70E-32	[Triticum aestivum]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
557	G865	AW560968	1.10E-28	[Medicago truncatula]	EST316016 DSIR Medicago truncatula cDNA
557	G865	AW782252	1.20E-26	[Glycine max]	sm03d11.y1 Gm-c1027 Glycine max cDNA clone GENO
557	G865	BI421895	3.60E-25	[Lycopersicon esculentum]	EST532561 tomato callus, TAMU Lycop
557	G865	BE642320	1.60E-24	[Ceratopteris richardii]	Cri2_5_L17_SP6 Ceratopteris Spore Li
557	G865	BE494041	1.60E-24	[Secale cereale]	WHE1277_B09_D17ZS Secale cereale anther cDNA
557	G865	D39914	2.60E-24	[Oryza sativa]	RICS1576A Rice shoot Oryza sativa cDNA, mRNA s
557	G865	AV428124	9.00E-23	[Lotus japonicus]	AV428124 Lotus japonicus young plants (two- Tobacco mRNA for EREBP-2, complete cds.
557	G865	TOBBY4D	1.80E-21	[Nicotiana tabacum]	ERF1.
557	G865	gi1208495	2.40E-23	[Nicotiana tabacum]	ethylene-responsive element binding
557	G865	gi8809571	5.10E-23	[Nicotiana sylvestris]	Pt14.
557	G865	gi3342211	1.40E-22	[Lycopersicon esculentum]	AP2-related transcription f
557	G865	gi7528276	1.70E-22	[Mesembryanthemum crystallinum]	Putative AP2 domain containing protein.
557	G865	gi15217291	7.80E-22	[Oryza sativa]	AP2 domain containing protein.
557	G865	gi3264767	2.70E-21	[Prunus armeniaca]	AP2 domain DNA-binding protein.
557	G865	gi8980313	2.10E-20	[Catharanthus roseus]	apetala2 domain-containing protein.
557	G865	gi8571476	9.30E-20	[Atriplex hortensis]	DNA binding protein homolog.
557	G865	gi1688233	1.40E-19	[Solanum tuberosum]	ethylene-responsive element binding
557	G865	gi6478845	1.80E-19	[Matricaria chamomilla]	BOHOJ67TR BOHO Brassica oleracea genomic
23	G2509	BH577856	2.50E-29	[Brassica oleracea]	sak01e08.y1 Gm-c1074 Glycine max cDNA clone SOY
23	G2509	BM269574	5.90E-28	[Glycine max]	WWS012.C2R000101 ITEC WWS Wheat Scutellum
23	G2509	BE419451	2.20E-27	[Triticum aestivum]	EST249507 tomato ovary, TAMU Lycop
23	G2509	AI483636	7.80E-27	[Lycopersicon esculentum]	EST316016 DSIR Medicago truncatula cDNA
23	G2509	AW560968	8.90E-27	[Medicago truncatula]	Cri2_5_L17_SP6 Ceratopteris Spore Li
23	G2509	BE642320	4.30E-26	[Ceratopteris richardii]	chromosome 1 clone P0677H08, *** SEQUENCING IN
23	G2509	AP003286	1.00E-25	[Oryza sativa]	WHE1277_B09_D17ZS Secale cereale anther cDNA
23	G2509	BE494041	3.20E-25	[Secale cereale]	HVSMEH0102106f Hordeum vulgare 5-45 DAP spi
23	G2509	BE602106	1.10E-24	[Hordeum vulgare]	AV428124 Lotus japonicus young plants (two- AP2 domain containing protein.
23	G2509	AV428124	1.00E-23	[Lotus japonicus]	
23	G2509	gi3264767	4.00E-27	[Prunus armeniaca]	

Table 5

23	G2509	gi12003376	1.40E-23	[Nicotiana tabacum]	Avr9/Cf-9 rapidly elicited protein 1.
23	G2509	gi14140141	2.30E-23	[Oryza sativa]	putative AP2-related transcription factor.
23	G2509	gi1688233	5.40E-23	[Solanum tuberosum]	DNA binding protein homolog.
23	G2509	gi4099921	2.60E-22	[Stylosanthes hamata]	EREBP-3 homolog.
23	G2509	gi8809571	7.80E-22	[Nicotiana glauca]	ethylene-responsive element binding
23	G2509	gi3342211	1.00E-21	[Lycopersicon esculentum]	Pti4.
23	G2509	gi7528276	2.70E-21	[Mesembryanthemum crystallinum]	AP2-related transcription f
23	G2509	gi17385636	1.90E-20	[Matricaria chamomilla]	ethylene-responsive element binding
23	G2509	gi18496063	3.30E-20	[Fagus sylvatica]	ethylene responsive element binding prote
1119	G2347	BI931517	5.30E-31	[Lycopersicon esculentum]	EST551406 tomato flower, 8 mm to pr
1119	G2347	BE058432	4.20E-29	[Glycine max]	sn16a06.y1 Gm-c1016 Glycine max cDNA clone GENO
1119	G2347	AMSPB1	1.80E-28	[Antirrhinum majus]	A.majus mRNA for squamosa-promoter bindin
1119	G2347	BG525285	5.70E-28	[Stevia rebaudiana]	48-3 Stevia field grown leaf cDNA Stevia
1119	G2347	L38193	4.60E-27	[Brassica rapa]	BNAF1025E Mustard flower buds Brassica rapa c
1119	G2347	BG455868	6.40E-27	[Medicago truncatula]	NF068F05PL1F1045 Phosphate starved leaf
1119	G2347	BG097153	1.70E-24	[Solanum tuberosum]	EST461672 potato leaves and petioles Sola
1119	G2347	BF482644	1.60E-23	[Triticum aestivum]	WHE2301-2304 A21 A21ZS Wheat pre-anthesis
1119	G2347	AW747167	2.30E-23	[Sorghum bicolor]	WS1 66 F11.b1 A002 Water-stressed 1 (WS1) S
1119	G2347	BG442540	2.50E-23	[Gossypium arboreum]	GA_Ea0017G06f Gossypium arboreum 7-10 d
1119	G2347	gi1183864	1.50E-31	[Antirrhinum majus]	squamosa-promoter binding protein 2.
1119	G2347	gi5931786	3.40E-25	[Zea mays]	SBP-domain protein 5.
1119	G2347	gi8468036	1.40E-21	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
1119	G2347	gi9087308	6.60E-09	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
1119	G2347	gi7209500	0.83	[Brassica rapa]	S-locus pollen protein.
43	G988	CRU303349	3.10E-208	[Capsella rubella]	ORF1, ORF2, ORF3, ORF4, ORF5 and ORF6 (pa
43	G988	A84072	4.50E-86	[Lycopersicon esculentum]	Sequence 1 from Patent WO9846759.
43	G988	A84080	3.30E-85	[Solanum tuberosum]	Sequence 9 from Patent WO9846759.
43	G988	AP003944	1.30E-57	[Oryza sativa]	chromosome 6 clone OJ1126 F05, *** SEQUENCING
43	G988	AX081276	2.80E-43	[Brassica napus]	Sequence 1 from Patent WO0109356.
43	G988	ZMA242530	1.50E-40	[Zea mays]	partial d8 gene for gibberellin response modula
43	G988	AX005804	2.50E-37	[Triticum aestivum]	Sequence 13 from Patent WO9909174.
43	G988	AB048713	9.10E-33	[Pisum sativum]	PsSCR mRNA for SCARECROW, complete cds.
43	G988	AW774515	2.00E-29	[Medicago truncatula]	EST333666 KV3 Medicago truncatula cDNA
43	G988	BE822458	1.20E-27	[Glycine max]	GM700017A20H12 Gm-1070 Glycine max cDNA clone
43	G988	gi13620166	8.00E-211	[Capsella rubella]	hypothetical protein.
43	G988	gi4160441	1.40E-87	[Lycopersicon esculentum]	lateral suppressor protein.

Table 5

43	G988	gi10178637	2.20E-48	[Zea mays]	SCARECROW.
43	G988	gi6970472	1.20E-47	[Oryza sativa]	OsGAL.
43	G988	gi5640157	2.80E-45	[Triticum aestivum]	gibberellin response modulator.
43	G988	gi13170126	7.10E-45	[Brassica napus]	unnamed protein product.
43	G988	gi13365610	1.10E-40	[Pisum sativum]	SCARECROW.
43	G988	gi14318115	1.10E-14	[Zea mays subsp. mays]	gibberellin response modulator.
43	G988	gi14318165	7.30E-14	[Triticum dactyloides]	gibberellin response modulator.
43	G988	gi347457	2.40E-05	[Glycine max]	hydroxyproline-rich glycoprotein.
459	G2346	AMA011622	3.10E-35	[Antirrhinum majus]	mRNA for squamosa promoter binding
459	G2346	AW691786	1.80E-26	[Medicago truncatula]	NF044B06ST1F1000 Developing stem Medica
459	G2346	AQ273505	7.00E-25	[Oryza sativa]	nbxb0030003f CUGI Rice BAC Library Oryza sativ
459	G2346	AW932595	7.90E-24	[Lycopersicon esculentum]	EST358438 tomato fruit mature green
459	G2346	BG593787	9.50E-24	[Solanum tuberosum]	EST492465 cSTS Solanum tuberosum cDNA clo
459	G2346	BG442540	1.00E-23	[Gossypium arboreum]	GA Ea0017G06f Gossypium arboreum 7-10 d
459	G2346	AZ919034	1.90E-23	[Zea mays]	1006013G02.x3 1006 - RescueMu Grid G Zea mays geno
459	G2346	BE596165	2.70E-23	[Sorghum bicolor]	PI1 50 D04.b1 A002 Pathogen induced 1 (PI1)
459	G2346	AI443033	2.30E-22	[Glycine max]	sa31a08.y1 Gm-c1004 Glycine max cDNA clone GENO
459	G2346	BF482644	4.30E-22	[Triticum aestivum]	WHE2301-2304 A21 A21ZS Wheat pre-anthesis
459	G2346	gi5931643	6.20E-45	[Antirrhinum majus]	squamosa promoter binding protein-homol
459	G2346	gi5931786	4.20E-26	[Zea mays]	SBP-domain protein 5.
459	G2346	gi8468036	3.30E-14	[Oryza sativa]	Similar to Arabidopsis thaliana chromosome 2
459	G2346	gi9087308	8.30E-08	[Mitochondrion Beta vulgaris var. altissima]	orf102a.
285	G1354	BG128374	2.90E-58	[Lycopersicon esculentum]	EST474020 tomato shoot/meristem Lyc
285	G1354	BE202831	1.90E-56	[Medicago truncatula]	EST402853 KV1 Medicago truncatula cDNA
285	G1354	AI161918	6.60E-55	[Populus tremula x Populus tremuloides]	A009P50U Hybrid aspen
285	G1354	AB028186	1.20E-53	[Oryza sativa]	mRNA for OsNAC7 protein, complete cds.
285	G1354	BE060921	8.00E-50	[Hordeum vulgare]	HVSMEg0013N15f Hordeum vulgare pre-anthesis
285	G1354	AF402603	1.50E-42	[Phaseolus vulgaris]	NAC domain protein NAC2 mRNA, complete c
285	G1354	BE357920	1.60E-42	[Sorghum bicolor]	DG1 23 F03.b1 A002 Dark Grown 1 (DG1) Sorgh
285	G1354	PHRNANAM	3.60E-42	[Petunia x hybrida]	P.hybrida mRNA encoding NAM protein.
285	G1354	AW185617	5.30E-40	[Glycine max]	se80b05.y1 Gm-c1023 Glycine max cDNA clone GENO
285	G1354	gi6006373	4.50E-63	[Oryza sativa]	Similar to NAM like protein (AC005310).
285	G1354	gi15148914	2.30E-44	[Phaseolus vulgaris]	NAC domain protein NAC2.
285	G1354	gi14485513	3.50E-44	[Solanum tuberosum]	putative NAC domain protein.
285	G1354	gi1279640	5.90E-44	[Petunia x hybrida]	NAM.
285	G1354	gi6175246	5.20E-41	[Lycopersicon esculentum]	jasmonic acid 2.

Table 5

285	G1354	gi4218535	5.10E-39	[Triticum sp.]	GRAB1 protein.
285	G1354	gi6732158	5.10E-39	[Triticum monococcum]	unnamed protein product.
285	G1354	gi7716952	3.30E-35	[Medicago truncatula]	NAC1.
285	G1354	gi4996349	2.50E-26	[Nicotiana tabacum]	NAC-domain protein.
285	G1354	gi2982275	3.10E-14	[Picea mariana]	ATAF1-like protein.
119	G1063	BH700922	4.50E-90	[Brassica oleracea]	BOMMZ07TR BO 2 3 KB Brassica oleracea gen
119	G1063	BE451174	2.40E-41	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
119	G1063	AW832545	2.00E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
119	G1063	AP004693	5.90E-37	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
119	G1063	AP004462	4.40E-32	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
119	G1063	AT002234	8.90E-32	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
119	G1063	BF263465	5.40E-25	[Hordeum vulgare]	HV_CEA0006N02f Hordeum vulgare seedling gre
119	G1063	BG557011	4.20E-22	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
119	G1063	BG842856	3.10E-21	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
119	G1063	BG559930	1.40E-18	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
119	G1063	gi15528743	4.20E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
119	G1063	gi6166283	8.10E-10	[Pinus taeda]	helix-loop-helix protein 1A.
119	G1063	gi11045087	8.80E-09	[Brassica napus]	putative protein.
119	G1063	gi10998404	7.10E-08	[Petunia x hybrida]	anthocyanin 1.
119	G1063	gi99441	2.60E-07	[Volvox carter]	sulfated surface glycoprotein 185 - Volvox
119	G1063	gi1142621	5.00E-07	[Phaseolus vulgaris]	phaseolin G-box binding protein PG2.
119	G1063	gi166428	8.10E-07	[Antirrhinum majus]	DEL.
119	G1063	gi1247386	9.50E-07	[Nicotiana glauca]	PRP2.
119	G1063	gi82091	1.00E-06	[Lycopersicon esculentum]	hydroxyproline-rich glycoprotein
119	G1063	gi7486263	1.40E-06	[Catharanthus roseus]	extensin.
129	G2143	BH650724	3.00E-88	[Brassica oleracea]	BOMIW43TR BO 2 3 KB Brassica oleracea gen
129	G2143	AW832545	1.50E-40	[Glycine max]	sm12e10.y1 Gm-c1027 Glycine max cDNA clone GENO
129	G2143	BE451174	3.50E-40	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
129	G2143	AP004693	4.00E-38	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
129	G2143	AP004584	6.30E-33	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
129	G2143	AT002234	3.00E-31	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
129	G2143	BF263465	2.90E-26	[Hordeum vulgare]	HV_CEA0006N02f Hordeum vulgare seedling gre
129	G2143	BG557011	2.60E-22	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
129	G2143	BG842856	3.50E-20	[Zea mays]	MEST40-H05.T3 ISUM4-TN Zea mays cDNA clone MEST40-
129	G2143	BG559930	6.10E-18	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
129	G2143	gi15528743	5.50E-26	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.

Table 5

129	G2143	gi1086538	7.60E-09	[Oryza rufipogon]	transcriptional activator Rb homolog.
129	G2143	gi6166283	1.10E-08	[Pinus taeda]	helix-loop-helix protein 1A.
129	G2143	gi1142621	4.60E-07	[Phaseolus vulgaris]	phaseolin G-box binding protein PG2.
129	G2143	gi3399777	5.20E-07	[Glycine max]	symbiotic ammonium transporter; nodulin.
129	G2143	gi5923912	6.10E-07	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
129	G2143	gi10998404	9.20E-07	[Petunia x hybrida]	anthocyanin 1.
129	G2143	gi4321762	5.20E-06	[Zea mays]	transcription factor MYC7E.
129	G2143	gi166428	6.00E-06	[Antirrhinum majus]	DEL.
129	G2143	gi527665	7.40E-06	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	BH511840	6.70E-62	[Brassica oleracea]	BOGRJ19TR BOGR Brassica oleracea genomic
133	G2557	BE347811	3.70E-46	[Glycine max]	sp05h10.y1 Gm-c1041 Glycine max cDNA clone GENO
133	G2557	AP003141	2.40E-33	[Oryza sativa]	genomic DNA, chromosome 1, PAC clone:P0002B05.
133	G2557	BF263465	3.00E-31	[Hordeum vulgare]	HV CEa0006N02f Hordeum vulgare seedling gre
133	G2557	AT002234	6.60E-27	[Brassica rapa subsp. pekinensis]	AT002234 Flower bud cDNA Br
133	G2557	BG557011	6.40E-26	[Sorghum bicolor]	EM1_41_E02.g1_A002 Embryo 1 (EM1) Sorghum b
133	G2557	AP004462	7.90E-26	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
133	G2557	BE451174	3.90E-25	[Lycopersicon esculentum]	EST402062 tomato root, plants pre-a
133	G2557	BG842856	5.60E-22	[Zea mays]	MEST40-H05.T3 SUM4-TN Zea mays cDNA clone MEST40-
133	G2557	BG559930	7.00E-14	[Sorghum propinquum]	RHIZ2_75_D09.g1_A003 Rhizome2 (RHIZ2) So
133	G2557	gi15289790	2.40E-36	[Oryza sativa]	contains EST C74560(E31855)-unknown protein.
133	G2557	gi3399777	2.60E-06	[Glycine max]	symbiotic ammonium transporter; nodulin.
133	G2557	gi4206118	1.10E-05	[Mesembryanthemum crystallinum]	transporter homolog.
133	G2557	gi6166283	1.30E-05	[Pinus taeda]	helix-loop-helix protein 1A.
133	G2557	gi527655	3.70E-05	[Pennisetum glaucum]	myc-like regulatory R gene product.
133	G2557	gi5923912	3.70E-05	[Tulipa gesneriana]	bHLH transcription factor GBOF-1.
133	G2557	gi527661	7.80E-05	[Phyllostachys acuta]	myc-like regulatory R gene product.
133	G2557	gi527665	9.50E-05	[Sorghum bicolor]	myc-like regulatory R gene product.
133	G2557	gi1086538	0.0001	[Oryza rufipogon]	transcriptional activator Rb homolog.
133	G2557	gi5669656	0.00013	[Lycopersicon esculentum]	ER33 protein.
697	G2430	BF632520	1.90E-14	[Medicago truncatula]	NF039A08DT1F1054 Drought Medicago trunc
697	G2430	AW396912	1.20E-13	[Glycine max]	sg64g09.y1 Gm-c1007 Glycine max cDNA clone GENO
697	G2430	D41804	4.50E-13	[Oryza sativa]	RICS4626A Rice shoot Oryza sativa cDNA, mRNA s
697	G2430	BE214029	2.60E-10	[Hordeum vulgare]	HV CEb0001P06f Hordeum vulgare seedling gre
697	G2430	AW564570	2.70E-10	[Sorghum bicolor]	LG1_296_E01.b1_A002 Light Grown 1 (LG1) Sor
697	G2430	BG129795	5.40E-10	[Lycopersicon esculentum]	EST475441 tomato shoot/meristem Lyc
697	G2430	AB060130	5.40E-09	[Zea mays]	ZmRRR8 mRNA for response regulator 8, complete cds.

Table 5

697	G2430	BF587105	2.50E-05	[Sorghum propinquum]	FM1_32_C05.b1_A003 Floral-Induced Merist
697	G2430	AI163121	0.3	[Populus tremula x Populus tremuloides]	A033P70U Hybrid aspen
697	G2430	BG595628	0.46	[Solanum tuberosum]	EST494306 cSTS Solanum tuberosum cDNA clo
697	G2430	gi13661174	5.40E-18	[Zea mays]	response regulator 8.
697	G2430	gi15289981	0.028	[Oryza sativa]	hypothetical protein.
697	G2430	gi6942190	0.12	[Mesembryanthemum crystallinum]	CDPK substrate protein 1; C
697	G2430	gi4519671	0.2	[Nicotiana tabacum]	transfactor:
831	G1478	BF275913	1.50E-20	[Gossypium arboreum]	GA_Eb0025C07f Gossypium arboreum 7-10 d
831	G1478	BG157399	6.50E-19	[Glycine max]	sab36g12.y1 Gm-c1026 Glycine max cDNA clone GEN
831	G1478	C95300	2.20E-10	[Citrus unshiu]	C95300 Citrus unshiu Miyagawa-wase maturation
831	G1478	AW034552	2.70E-10	[Lycopersicon esculentum]	EST278168 tomato callus, TAMU Lycop
831	G1478	BI070429	3.40E-10	[Populus tremula x Populus tremuloides]	C037P68U Populus tra
831	G1478	AF016011	5.10E-09	[Brassica napus]	CONSTANS homolog (Bn9CON10) gene, complete c
831	G1478	BE598912	6.20E-09	[Sorghum bicolor]	PI1_84_H11.b1_A002 Pathogen induced 1 (P11)
831	G1478	BG605313	6.80E-09	[Triticum aestivum]	WHE2331_C04_F07ZS Wheat pre-anthesis spik
831	G1478	BE558327	8.90E-09	[Hordeum vulgare]	HV_CE0017D19f Hordeum vulgare seedling gre
831	G1478	BG647091	1.20E-08	[Medicago truncatula]	EST508710 HOGA Medicago truncatula cDNA
831	G1478	gi2895188	4.70E-11	[Brassica napus]	CONSTANS homolog.
831	G1478	gi3618308	1.50E-09	[Oryza sativa]	zinc finger protein.
831	G1478	gi11037308	4.70E-09	[Brassica nigra]	constans-like protein.
831	G1478	gi3341723	1.30E-08	[Raphanus sativus]	CONSTANS-like 1 protein.
831	G1478	gi4091806	1.50E-07	[Malus x domestica]	CONSTANS-like protein 2.
831	G1478	gi10946337	3.10E-07	[Ipomoea nil]	CONSTANS-like protein.
831	G1478	gi4557093	1.40E-05	[Pinus radiata]	zinc finger protein.
831	G1478	gi619312	0.9	[Capparis masakal]	mabinlin III B-chain=sweet protein mabi
831	G1478	gi4732091	1	[Zea mays]	bundle sheath defective protein 2.
831	G1478	gi4699629	1	[Nicotiana glauca]	Chain A, Putative Ancestral Protein Encod
579	G681	BG128147	6.80E-41	[Lycopersicon esculentum]	EST473793 tomato shoot/meristem Lyc
579	G681	BF054497	1.50E-39	[Solanum tuberosum]	EST439727 potato leaves and petioles Sola
579	G681	BE054276	8.40E-39	[Gossypium arboreum]	GA_Ea0002O18f Gossypium arboreum 7-10 d
579	G681	BG269414	4.00E-38	[Mesembryanthemum crystallinum]	LO-3478T3 Ice plant Lambda Un
579	G681	BF620286	7.40E-38	[Hordeum vulgare]	HVSMec0019F08f Hordeum vulgare seedling sho
579	G681	BE490032	1.00E-37	[Triticum aestivum]	WHE0364_C04_E08ZS Wheat cold-stressed see
579	G681	BI542536	1.40E-36	[Zea mays]	949021A03.y1 949 - Juvenile leaf and shoot cDNA fr
579	G681	BF425254	7.20E-36	[Glycine max]	su42c10.y1 Gm-c1068 Glycine max cDNA clone GENO
579	G681	AW672062	3.20E-34	[Sorghum bicolor]	LG1_354_G05.b1_A002 Light Grown 1 (LG1) Sor

Table 5

579	G681	BG448527	1.00E-33	[Medicago truncatula]	NF036F04RT1F1032 Developing root Medica
579	G681	gi13346188	9.10E-37	[Gossypium hirsutum]	GhMYB25.
579	G681	gi20563	6.30E-36	[Petunia x hybrida]	protein 1.
579	G681	gi485867	1.20E-34	[Antirrhinum majus]	mixta.
579	G681	gi2605617	1.70E-32	[Oryza sativa]	OSMYB1.
579	G681	gi1430846	2.00E-31	[Lycopersicon esculentum]	myb-related transcription factor.
579	G681	gi6651292	2.20E-30	[Pimpinella brachycarpa]	myb-related transcription factor.
579	G681	gi15042116	4.90E-30	[Zea mays subsp. parviglumis]	Cl protein.
579	G681	gi82730	6.10E-30	[Zea mays]	transforming protein (myb) homolog (clone Zm38)
579	G681	gi5139806	8.30E-30	[Glycine max]	GmMYB29A2.
579	G681	gi19055	1.10E-29	[Hordeum vulgare]	MybHV5.
611	G878	AF096299	6.20E-90	[Nicotiana tabacum]	DNA-binding protein 2 (WRKY2) mRNA, compl
611	G878	CUSSLD	1.80E-83	[Cucumis sativus]	SPF1-like DNA-binding protein mRNA, complet
611	G878	AF193802	3.50E-63	[Oryza sativa]	zinc finger transcription factor WRKY1 mRNA, c
611	G878	AX192162	2.20E-62	[Glycine max]	Sequence 9 from Patent WO0149840.
611	G878	IPBSPF1P	3.80E-58	[Ipomoea batatas]	Sweet potato mRNA for SPF1 protein, complet
611	G878	AFABF1	2.00E-56	[Avena fatua]	A. fatua mRNA for DNA-binding protein (clone ABF
611	G878	LES303343	7.20E-55	[Lycopersicon esculentum]	mRNA for hypothetical protein (ORF
611	G878	AX192164	4.00E-54	[Triticum aestivum]	Sequence 11 from Patent WO0149840.
611	G878	AF080595	2.10E-53	[Pimpinella brachycarpa]	zinc finger protein (ZFP1) mRNA, com
611	G878	PCU48831	2.30E-53	[Petroselinum crispum]	DNA-binding protein WRKY1 mRNA, comple
611	G878	gi4322940	3.30E-128	[Nicotiana tabacum]	DNA-binding protein 2.
611	G878	gi927025	1.10E-109	[Cucumis sativus]	SPF1-like DNA-binding protein.
611	G878	gi6689916	1.50E-74	[Oryza sativa]	zinc finger transcription factor WRKY1.
611	G878	gi484261	1.10E-66	[Ipomoea batatas]	SPF1 protein.
611	G878	gi1159877	2.30E-63	[Avena fatua]	DNA-binding protein.
611	G878	gi13620227	4.60E-63	[Lycopersicon esculentum]	hypothetical protein.
611	G878	gi5917653	1.70E-56	[Petroselinum crispum]	zinc-finger type transcription facto
611	G878	gi4894965	5.00E-56	[Avena sativa]	DNA-binding protein WRKY1.
611	G878	gi3420906	8.70E-56	[Pimpinella brachycarpa]	zinc finger protein; WRKY1.
611	G878	gi13620168	4.20E-22	[Capsella rubella]	hypothetical protein.
47	G374	AP004457	1.20E-73	[Oryza sativa (japonica cultivar-group)]	() chromosome 8 clo
47	G374	AP004693	1.90E-73	[Oryza sativa]	chromosome 8 clone P0461F06, *** SEQUENCING IN
47	G374	BH552835	1.30E-62	[Brassica oleracea]	BOHG567R BOHG Brassica oleracea genomic
47	G374	BG128229	6.50E-55	[Lycopersicon esculentum]	EST473875 tomato shoot/meristem Lyc
47	G374	BG646959	3.20E-46	[Medicago truncatula]	EST508578 HOGA Medicago truncatula cDNA

Table 5

47	G374	BG890162	8.70E-41	[Solanum tuberosum]	EST516013 cSTD Solanum tuberosum cDNA clo
47	G374	AW179366	6.00E-38	[Zea mays]	618046G06.y1 618 - Inbred Tassel cDNA Library Zea
47	G374	BF473206	1.50E-32	[Triticum aestivum]	WHE0922_G12_M24ZS Wheat 5-15 DAP spike cD
47	G374	AW761011	2.90E-29	[Glycine max]	sl61g11.y1 Gm-c1027 Glycine max cDNA clone GENO
47	G374	AJ436050	1.50E-27	[Hordeum vulgare]	AJ436050 S00007 Hordeum vulgare cDNA clone
47	G374	gi422012	0.8	[Sorghum bicolor]	lipid transfer protein - sorghum (fragmen
47	G374	gi1827893	1	[Zea mays]	Maize Nonspecific Lipid Transfer Protein Complex

Traits of interest

Examples of some of the traits that may be desirable in plants, and that may be provided by transforming the plants with the presently disclosed sequences, are listed in Table 6.

Table 6. Genes, traits and utilities that affect plant characteristics

<u>Trait Category</u>	<u>Traits</u>	Transcription factor genes that <u>impact traits</u>	<u>Utility</u> <u>Gene effect on:</u>
Resistance and tolerance	Salt stress resistance	G22; G196; G226; G303; G312; G325; G353; G482; G545; G801; G867; G884; G922; G926; G1452; G1794; G1820; G1836; G1843; G1863; G2053; G2110; G2140; G2153; G2379; G2701; G2713; G2719; G2789	Germination rate, survivability, yield; extended growth range
	Osmotic stress resistance	G47; G175; G188; G303; G325; G353; G489; G502; G526; G921; G922; G926; G1069; G1089; G1452; G1794; G1930; G2140; G2153; G2379; G2701; G2719; G2789;	Germination rate, survivability, yield
	Cold stress resistance; cold germination	G256; G394; G664; G864; G1322; G2130	Germination, growth, earlier planting
	Tolerance to freezing	G303; G325; G353; G720; G912; G913; G1794; G2053; G2140; G2153; G2379; G2701; G2719; G2789	Survivability, yield, appearance, extended range
	Heat stress resistance	G3; G464; G682; G864; G964;	Germination,

		G1305; G1645; G2130 G2430	growth, later planting
	Drought, low humidity resistance	G303; G325; G353; G720; G912; G926; G1452; G1794; G1820; G1843; G2053; G2140; G2153; G2379; G2583; G2701; G2719; G2789	Survivability, yield, extended range
	Radiation resistance	G1052	Survivability, vigor, appearance
	Decreased herbicide sensitivity	G343; G2133; G2517	Resistant to increased herbicide use
	Increased herbicide sensitivity	G374; G877; G1519	Use as a herbicide target
	Oxidative stress	G477; G789; G1807; G2133; G2517	Improved yield, appearance, reduced senescence
	Light response	G183; G354; G375; G1062; G1322; G1331; G1488; G1494; G1521; G1786; G1794; G2144; G2555;	Germination, growth, development, flowering time
Development, morphology	Overall plant architecture	G24; G27; G31; G33; G47; G147; G156; G160; G182; G187; G195; G196; G211; G221; G237; G280; G342; G352; G357; G358; G360; G362; G364; G365; G367; G373; G377; G396; G431; G447; G479; G546; G546; G551; G578; G580; G596; G615; G617; G620; G625;	Vascular tissues, lignin content; cell wall content; appearance

		G638; G658; G716; G725; G727; G730; G740; G770; G858; G865; G869; G872; G904; G910; G912; G920; G939; G963; G977; G979; G987; G988; G993; G1007; G1010; G1014; G1035; G1046; G1049; G1062; G1069; G1070; G1076; G1089; G1093; G1127; G1131; G1145; G1229; G1246; G1304; G1318; G1320; G1330; G1331; G1352; G1354; G1360; G1364; G1379; G1384; G1399; G1415; G1417; G1442; G1453; G1454; G1459; G1460; G1471; G1475; G1477; G1487; G1487; G1492; G1499; G1499; G1531; G1540; G1543; G1543; G1544; G1548; G1584; G1587; G1588; G1589; G1636; G1642; G1747; G1749; G1749; G1751; G1752; G1763; G1766; G1767; G1778; G1789; G1790; G1791; G1793; G1794; G1795; G1800; G1806; G1811; G1835; G1836; G1838; G1839; G1843; G1853; G1855; G1865; G1881; G1882; G1883; G1884; G1891; G1896; G1898; G1902; G1904; G1906; G1913; G1914; G1925; G1929; G1930; G1954; G1958; G1965; G1976; G2057; G2107; G2133; G2134; G2151; G2154; G2157; G2181;	
--	--	--	--

		G2290; G2299; G2340; G2340; G2346; G2373; G2376; G2424; G2465; G2505; G2509; G2512; G2513; G2519; G2520; G2533; G2534; G2573; G2589; G2687; G2720; G2787; G2789; G2893	
	Size: increased stature	G189; G1073; G1435; G2430	
	Size: reduced stature or dwarfism	G3; G5; G21; G23; G39; G165; G184; G194; G258; G280; G340; G343; G353; G354; G362; G363; G370; G385; G396; G439; G440; G447; G450; G550; G557; G599; G636; G652; G670; G671; G674; G729; G760; G804; G831; G864; G884; G898; G900; G912; G913; G922; G932; G937; G939; G960; G962; G977; G991; G1000; G1008; G1020; G1023; G1053; G1067; G1075; G1137; G1181; G1198; G1228; G1266; G1267; G1275; G1277; G1309; G1311; G1314; G1317; G1322; G1323; G1326; G1332; G1334; G1367; G1381; G1382; G1386; G1421; G1488; G1494; G1537; G1545; G1560; G1586; G1641; G1652; G1655; G1671; G1750; G1756; G1757; G1782; G1786; G1794; G1839; G1845; G1879; G1886; G1888; G1933; G1939; G1943; G1944; G2011; G2094; G2115;	Ornamental; small stature provides wind resistance; creation of dwarf varieties

		G2130; G2132; G2144; G2145; G2147; G2156; G2294; G2313; G2344; G2431; G2510; G2517; G2521; G2893; G2893	
	Fruit size and number	G362	Biomass, yield, cotton boll fiber density
	Flower structure, inflorescence	G47; G259; G353; G354; G671; G732; G988; G1000; G1063; G1140; G1326; G1449; G1543; G1560; G1587; G1645; G1947; G2108; G2143; G2893	Ornamental horticulture; production of saffron or other edible flowers
	Number and development of trichomes	G225; G226; G247; G362; G585; G634; G676; G682; G1014; G1332; G1452; G1795; G2105	Resistance to pests and desiccation; essential oil production
	Seed size, color, and number	G156; G450; G584; G652; G668; G858; G979; G1040; G1062; G1145; G1255; G1494; G1531; G1534; G1594; G2105; G2114;	Yield
	Root development, modifications	G9; G1482; G1534; G1794; G1852; G2053; G2136; G2140	
	Modifications to root hairs	G225; G226	Nutrient, water uptake, pathogen resistance
	Apical dominance	G559; G732; G1255; G1275; G1411; G1488; G1635; G2452; G2509	Ornamental horticulture
	Branching patterns	G568; G988; G1548	Ornamental horticulture, knot reduction, improved

			windscreen
	Leaf shape, color, modifications	G375; G377; G428; G438; G447; G464; G557; G577; G599; G635; G671; G674; G736; G804; G903; G977; G921; G922; G1038; G1063; G1067; G1073; G1075; G1146; G1152; G1198; G1267; G1269; G1452; G1484; G1586; G1594; G1767; G1786; G1792; G1886; G2059; G2094; G2105; G2113; G2117; G2143; G2144; G2431; G2452; G2465; G2587; G2583; G2724;	Appealing shape or shiny leaves for ornamental agriculture, increased biomass or photosynthesis
	Silique	G1134	Ornamental
	Stem morphology	G47; G438; G671; G748; G988; G1000	Ornamental; digestibility
	Shoot modifications	G390; G391	Ornamental stem bifurcations
Disease, Pathogen Resistance	Bacterial	G211; G347; G367; G418; G525; G545; G578; G1049	Yield, appearance, survivability, extended range
	Fungal	G19; G28; G28; G28; G147; G188; G207; G211; G237; G248; G278; G347; G367; G371; G378; G409; G477; G545; G545; G558; G569; G578; G591; G594; G616; G789; G805; G812; G865; G869; G872; G881; G896; G940; G1047; G1049; G1064; G1084; G1196; G1255; G1266;	Yield, appearance, survivability, extended range

		G1363; G1514; G1756; G1792; G1792; G1792; G1792; G1880; G1919; G1919; G1927; G1927; G1936; G1936; G1950; G2069; G2130; G2380; G2380; G2555	
Nutrients	Increased tolerance to nitrogen-limited soils	G225; G226; G1792	
	Increased tolerance to phosphate-limited soils	G419; G545; G561; G1946	
	Increased tolerance to potassium-limited soils	G561; G911	
Hormonal	Hormone sensitivity	G12; G546; G926; G760; G913; G926; G1062; G1069; G1095; G1134; G1330; G1452; G1666; G1820; G2140; G2789	Seed dormancy, drought tolerance; plant form, fruit ripening
Seed biochemistry	Production of seed prenyl lipids, including tocopherol	G214; G259; G490; G652; G748; G883; G1052; G1328; G1930; G2509; G2520	Antioxidant activity, vitamin E
	Production of seed sterols	G20	Precursors for human steroid hormones; cholesterol modulators
	Production of seed glucosinolates	G353; G484; G674; G1272; G1506; G1897; G1946; G2113; G2117; G2155; G2290; G2340	Defense against insects; putative anticancer activity; undesirable in

			animal feeds
	Modified seed oil content	G162; G162; G180; G192; G241; G265; G286; G291; G427; G509; G519; G561; G567; G590; G818; G849; G892; G961; G974; G1063; G1143; G1190; G1198; G1226; G1229; G1323; G1451; G1471; G1478; G1496; G1526; G1543; G1640; G1644; G1646; G1672; G1677; G1750; G1765; G1777; G1793; G1838; G1902; G1946; G1948; G2059; G2123; G2138; G2139; G2343; G2792; G2830	Vegetable oil production; increased caloric value for animal feeds; lutein content
	Modified seed oil composition	G217; G504; G622; G778; G791; G861; G869; G938; G965; G1417; G2192	Heat stability, digestibility of seed oils
	Modified seed protein content	G162; G226; G241; G371; G427; G509; G567; G597; G732; G849; G865; G892; G963; G988; G1323; G1323; G1419; G1478; G1488; G1634; G1637; G1641; G1644; G1652; G1677; G1777; G1777; G1818; G1820; G1903; G1909; G1946; G1946; G1958; G2059; G2117; G2417; G2509	Reduced caloric value for humans
Leaf biochemistry	Production of flavonoids	G1666*	Ornamental pigment production; pathogen resistance; health

			benefits
	Production of leaf glucosinolates	G264; G353; G484; G652; G674; G681; G1069; G1198; G1322; G1421; G1657; G1794; G1897; G1946; G2115; G2117; G2144; G2155; G2155; G2340; G2512; G2520; G2552	Defense against insects; putative anticancer activity; undesirable in animal feeds
	Production of diterpenes	G229	Induction of enzymes involved in alkaloid biosynthesis
	Production of anthocyanin	G546	Ornamental pigment
	Production of leaf phytosterols, inc. stigmastanol, campesterol	G561; G2131; G2424	Precursors for human steroid hormones; cholesterol modulators
	Leaf fatty acid composition	G214; G377; G861; G962; G975; G987; G1266; G1337; G1399; G1465; G1512; G2136; G2147; G2192	Nutritional value; increase in waxes for disease resistance
	Production of leaf prenyl lipids, including tocopherol	G214; G259; G280; G652; G987; G1543; G2509; G2520	Antioxidant activity, vitamin E
Biochemistry, general	Production of miscellaneous secondary metabolites	G229; G663	
	Sugar, starch, hemicellulose composition,	G158; G211; G211; G237; G242; G274; G598; G1012; G1266; G1309; G1309; G1641; G1765; G1865; G2094; G2094;	Food digestibility, hemicellulose & pectin content; fiber content; plant

		G2589; G2589	tensile strength, wood quality, pathogen resistance, pulp production; tuber starch content
Sugar sensing	Plant response to sugars	G26; G38; G43; G207; G218; G241; G254; G263; G308; G536; G567; G567; G680; G867; G912; G956; G996; G1068; G1225; G1314; G1314; G1337; G1759; G1804; G2153; G2379	Photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, senescence
Growth, Reproduction	Plant growth rate and development	G447; G617; G674; G730; G917; G937; G1035; G1046; G1131; G1425; G1452; G1459; G1492; G1589; G1652; G1879; G1943; G2430; G2431; G2465; G2521	Faster growth, increased biomass or yield, improved appearance; delay in bolting
	Embryo development	G167	
	Seed germination rate	G979; G1792; G2130	Yield
	Plant, seedling vigor	G561; G2346	Survivability, yield
	Senescence; cell death	G571; G636; G878; G1050; G1463; G1749; G1944; G2130; G2155; G2340; G2383	Yield, appearance; response to pathogens;
	Modified fertility	G39; G340; G439; G470; G559; G615; G652; G671; G779; G962; G977; G988; G1000; G1063; G1067; G1075;	Prevents or minimizes escape of the pollen of GMOs

		G1266; G1311; G1321; G1326; G1367; G1386; G1421; G1453; G1471; G1453; G1560; G1594; G1635; G1750; G1947; G2011; G2094; G2113; G2115; G2130; G2143; G2147; G2294; G2510; G2893	
	Early flowering	G147; G157; G180; G183; G183; G184; G185; G208; G227; G294; G390; G390; G390; G391; G391; G427; G427; G490; G565; G590; G592; G720; G789; G865; G898; G898; G989; G989; G1037; G1037; G1142; G1225; G1225; G1226; G1242; G1305; G1305; G1380; G1380; G1480; G1480; G1488; G1494; G1545; G1545; G1649; G1706; G1760; G1767; G1767; G1820; G1841; G1841; G1842; G1843; G1843; G1946; G1946; G2010; G2030; G2030; G2144; G2144; G2295; G2295; G2347; G2348; G2348; G2373; G2373; G2509; G2509; G2555; G2555	Faster generation time; synchrony of flowering; potential for introducing new traits to single variety
	Delayed flowering	G8; G47; G192; G214; G234; G361; G362; G562; G568; G571; G591; G680; G736; G748; G859; G878; G910; G912; G913; G971; G994; G1051; G1052; G1073; G1079; G1335; G1435; G1452; G1478;	Delayed time to pollen production of GMO plants; synchrony of flowering; increased yield

		G1789; G1804; G1865; G1865; G1895; G1900; G2007; G2133; G2155; G2291; G2465	
	Extended flowering phase	G1947	
	Flower and leaf development	G259; G353; G377; G580; G638 G652; G858; G869; G917; G922; G932; G1063; G1075; G1140; G1425; G1452; G1499; G1548; G1645; G1865; G1897; G1933; G2094; G2124; G2140; G2143; G2535; G2557	Ornamental applications; decreased fertility
	Flower abscission	G1897	Ornamental: longer retention of flowers

* When co-expressed with G669 and G663

Significance of modified plant traits

Currently, the existence of a series of maturity groups for different latitudes represents a major barrier to the introduction of new valuable traits. Any trait (e.g. disease resistance) has to be bred into each of the different maturity groups separately, a laborious and costly exercise. The availability of single strain, which could be grown at any latitude, would therefore greatly increase the potential for introducing new traits to crop species such as soybean and cotton.

For many of the traits, listed in Table 6 and below, that may be conferred to plants, a single transcription factor gene may be used to increase or decrease, advance or delay, or improve or prove deleterious to a given trait. For example, overexpression of a transcription factor gene that naturally occurs in a plant may cause early flowering relative to non-transformed or wild-type plants. By knocking out the gene, or suppressing the gene (with, for example, antisense suppression) the plant may experience delayed flowering. Similarly, overexpressing or suppressing one or more genes can impart significant differences in production of plant products,

such as different fatty acid ratios. Thus, suppressing a gene that causes a plant to be more sensitive to cold may improve a plant's tolerance of cold.

Salt stress resistance. Soil salinity is one of the more important variables that determines where a plant may thrive. Salinity is especially important for the successful cultivation of crop plants, particular in many parts of the world that have naturally high soil salt concentrations, or where the soil has been over-utilized. Thus, presently disclosed transcription factor genes that provide increased salt tolerance during germination, the seedling stage, and throughout a plant's life cycle would find particular value for imparting survivability and yield in areas where a particular crop would not normally prosper.

Osmotic stress resistance. Presently disclosed transcription factor genes that confer resistance to osmotic stress may increase germination rate under adverse conditions, which could impact survivability and yield of seeds and plants.

Cold stress resistance. The potential utility of presently disclosed transcription factor genes that increase tolerance to cold is to confer better germination and growth in cold conditions. The germination of many crops is very sensitive to cold temperatures. Genes that would allow germination and seedling vigor in the cold would have highly significant utility in allowing seeds to be planted earlier in the season with a high rate of survivability. Transcription factor genes that confer better survivability in cooler climates allow a grower to move up planting time in the spring and extend the growing season further into autumn for higher crop yields.

Tolerance to freezing. The presently disclosed transcription factor genes that impart tolerance to freezing conditions are useful for enhancing the survivability and appearance of plants conditions or conditions that would otherwise cause extensive cellular damage. Thus, germination of seeds and survival may take place at temperatures significantly below that of the mean temperature required for germination of seeds and survival of non-transformed plants. As with salt tolerance, this has the added benefit of increasing the potential range of a crop plant into regions in which it would otherwise succumb. Cold tolerant transformed plants may also be

planted earlier in the spring or later in autumn, with greater success than with non-transformed plants.

Heat stress tolerance. The germination of many crops is also sensitive to high temperatures. Presently disclosed transcription factor genes that provide increased heat tolerance are generally useful in producing plants that germinate and grow in hot conditions, may find particular use for crops that are planted late in the season, or extend the range of a plant by allowing growth in relatively hot climates.

Drought, low humidity tolerance. Strategies that allow plants to survive in low water conditions may include, for example, reduced surface area or surface oil or wax production. A number of presently disclosed transcription factor genes increase a plant's tolerance to low water conditions and provide the benefits of improved survivability, increased yield and an extended geographic and temporal planting range.

Radiation resistance. Presently disclosed transcription factor genes have been shown to increase lutein production. Lutein, like other xanthophylls such as zeaxanthin and violaxanthin, are important in the protection of plants against the damaging effects of excessive light. Lutein contributes, directly or indirectly, to the rapid rise of non-photochemical quenching in plants exposed to high light. Increased tolerance of field plants to visible and ultraviolet light impacts survivability and vigor, particularly for recent transplants. Also affected are the yield and appearance of harvested plants or plant parts. Crop plants engineered with presently disclosed transcription factor genes that cause the plant to produce higher levels of lutein therefore would have improved photoprotection, leading to less oxidative damage and increase vigor, survivability and higher yields under high light and ultraviolet light conditions.

Decreased herbicide sensitivity. Presently disclosed transcription factor genes that confer resistance or tolerance to herbicides (e.g., glyphosate) may find use in providing means to increase herbicide applications without detriment to desirable plants. This would allow for the increased use of a particular herbicide in a local

environment, with the effect of increased detriment to undesirable species and less harm to transgenic, desirable cultivars.

Increased herbicide sensitivity. Knockouts of a number of the presently disclosed transcription factor genes have been shown to be lethal to developing embryos. Thus, these genes are potentially useful as herbicide targets.

Oxidative stress. In plants, as in all living things, abiotic and biotic stresses induce the formation of oxygen radicals, including superoxide and peroxide radicals. This has the effect of accelerating senescence, particularly in leaves, with the resulting loss of yield and adverse effect on appearance. Generally, plants that have the highest level of defense mechanisms, such as, for example, polyunsaturated moieties of membrane lipids, are most likely to thrive under conditions that introduce oxidative stress (e.g., high light, ozone, water deficit, particularly in combination). Introduction of the presently disclosed transcription factor genes that increase the level of oxidative stress defense mechanisms would provide beneficial effects on the yield and appearance of plants. One specific oxidizing agent, ozone, has been shown to cause significant foliar injury, which impacts yield and appearance of crop and ornamental plants. In addition to reduced foliar injury that would be found in ozone resistant plant created by transforming plants with some of the presently disclosed transcription factor genes, the latter have also been shown to have increased chlorophyll fluorescence (Yu-Sen Chang et al. Bot. Bull. Acad. Sin. (2001) 42: 265-272).

Heavy metal tolerance. Heavy metals such as lead, mercury, arsenic, chromium and others may have a significant adverse impact on plant respiration. Plants that have been transformed with presently disclosed transcription factor genes that confer improved resistance to heavy metals, through, for example, sequestering or reduced uptake of the metals will show improved vigor and yield in soils with relatively high concentrations of these elements. Conversely, transgenic transcription factors may also be introduced into plants to confer an increase in heavy metal uptake, which may benefit efforts to clean up contaminated soils.

Light response. Presently disclosed transcription factor genes that modify a plant's response to light may be useful for modifying a plant's growth or

development, for example, photomorphogenesis in poor light, or accelerating flowering time in response to various light intensities, quality or duration to which a non-transformed plant would not similarly respond. Examples of such responses that have been demonstrated include leaf number and arrangement, and early flower bud appearances.

Overall plant architecture. Several presently disclosed transcription factor genes have been introduced into plants to alter numerous aspects of the plant's morphology. For example, it has been demonstrated that a number of transcription factors may be used to manipulate branching, such as the means to modify lateral branching, a possible application in the forestry industry. Transgenic plants have also been produced that have altered cell wall content, lignin production, flower organ number, or overall shape of the plants. Presently disclosed transcription factor genes transformed into plants may be used to affect plant morphology by increasing or decreasing internode distance, both of which may be advantageous under different circumstances. For example, for fast growth of woody plants to provide more biomass, or fewer knots, increased internode distances are generally desirable. For improved wind screening of shrubs or trees, or harvesting characteristics of, for example, members of the Gramineae family, decreased internode distance may be advantageous. These modifications would also prove useful in the ornamental horticulture industry for the creation of unique phenotypic characteristics of ornamental plants.

Increased stature. For some ornamental plants, the ability to provide larger varieties may be highly desirable. For many plants, including fruit-bearing trees or trees and shrubs that serve as view or wind screens, increased stature provides obvious benefits. Crop species may also produce higher yields on larger cultivars.

Reduced stature or dwarfism. Presently disclosed transcription factor genes that decrease plant stature can be used to produce plants that are more resistant to damage by wind and rain, or more resistant to heat or low humidity or water deficit. Dwarf plants are also of significant interest to the ornamental horticulture industry, and particularly for home garden applications for which space availability may be limited.

Fruit size and number. Introduction of presently disclosed transcription factor genes that affect fruit size will have desirable impacts on fruit size and number, which may comprise increases in yield for fruit crops, or reduced fruit yield, such as when vegetative growth is preferred (e.g., with bushy ornamentals, or where fruit is undesirable, as with ornamental olive trees).

Flower structure, inflorescence, and development. Presently disclosed transgenic transcription factors have been used to create plants with larger flowers or arrangements of flowers that are distinct from wild-type or non-transformed cultivars. This would likely have the most value for the ornamental horticulture industry, where larger flowers or interesting presentations generally are preferred and command the highest prices. Flower structure may have advantageous effects on fertility, and could be used, for example, to decrease fertility by the absence, reduction or screening of reproductive components. One interesting application for manipulation of flower structure, for example, by introduced transcription factors could be in the increased production of edible flowers or flower parts, including saffron, which is derived from the stigmas of *Crocus sativus*.

Number and development of trichomes. Several presently disclosed transcription factor genes have been used to modify trichome number and amount of trichome products in plants. Trichome glands on the surface of many higher plants produce and secrete exudates that give protection from the elements and pests such as insects, microbes and herbivores. These exudates may physically immobilize insects and spores, may be insecticidal or ant-microbial or they may act as allergens or irritants to protect against herbivores. Trichomes have also been suggested to decrease transpiration by decreasing leaf surface air flow, and by exuding chemicals that protect the leaf from the sun.

Seed size, color and number. The introduction of presently disclosed transcription factor genes into plants that alter the size or number of seeds may have a significant impact on yield, both when the product is the seed itself, or when biomass of the vegetative portion of the plant is increased by reducing seed production. In the case of fruit products, it is often advantageous to modify a plant to have reduced size

or number of seeds relative to non-transformed plants to provide seedless or varieties with reduced numbers or smaller seeds. Presently disclosed transcription factor genes have also been shown to affect seed size, including the development of larger seeds. Seed size, in addition to seed coat integrity, thickness and permeability, seed water content and by a number of other components including antioxidants and oligosaccharides, may affect seed longevity in storage. This would be an important utility when the seed of a plant is the harvested crops, as with, for example, peas, beans, nuts, etc. Presently disclosed transcription factor genes have also been used to modify seed color, which could provide added appeal to a seed product.

Root development, modifications. By modifying the structure or development of roots by transforming into a plant one or more of the presently disclosed transcription factor genes, plants may be produced that have the capacity to thrive in otherwise unproductive soils. For example, grape roots that extend further into rocky soils, or that remain viable in waterlogged soils, would increase the effective planting range of the crop. It may be advantageous to manipulate a plant to produce short roots, as when a soil in which the plant will be growing is occasionally flooded, or when pathogenic fungi or disease-causing nematodes are prevalent.

Modifications to root hairs. Presently disclosed transcription factor genes that increase root hair length or number potentially could be used to increase root growth or vigor, which might in turn allow better plant growth under adverse conditions such as limited nutrient or water availability.

Apical dominance. The modified expression of presently disclosed transcription factors that control apical dominance could be used in ornamental horticulture, for example, to modify plant architecture.

Branching patterns. Several presently disclosed transcription factor genes have been used to manipulate branching, which could provide benefits in the forestry industry. For example, reduction in the formation of lateral branches could reduce knot formation. Conversely, increasing the number of lateral branches could provide utility when a plant is used as a windscreen, or may also provide ornamental advantages.

Leaf shape, color and modifications. It has been demonstrated in laboratory experiments that overexpression of some of the presently disclosed transcription factors produced marked effects on leaf development. At early stages of growth, these transgenic seedlings developed narrow, upward pointing leaves with long petioles, possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements. Other transcription factor genes can be used to increase plant biomass; large size would be useful in crops where the vegetative portion of the plant is the marketable portion.

Siliques. Genes that later silique conformation in brassicates may be used to modify fruit ripening processes in brassicates and other plants, which may positively affect seed or fruit quality.

Stem morphology and shoot modifications. Laboratory studies have demonstrated that introducing several of the presently disclosed transcription factor genes into plants can cause stem bifurcations in shoots, in which the shoot meristems split to form two or three separate shoots. This unique appearance would be desirable in ornamental applications.

Diseases, pathogens and pests. A number of the presently disclosed transcription factor genes have been shown to or are likely to confer resistance to various plant diseases, pathogens and pests. The offending organisms include fungal pathogens *Fusarium oxysporum*, *Botrytis cinerea*, *Sclerotinia sclerotiorum*, and *Erysiphe orontii*. Bacterial pathogens to which resistance may be conferred include *Pseudomonas syringae*. Other problem organisms may potentially include nematodes, mollicutes, parasites, or herbivorous arthropods. In each case, one or more transformed transcription factor genes may provide some benefit to the plant to help prevent or overcome infestation. The mechanisms by which the transcription factors work could include increasing surface waxes or oils, surface thickness, local senescence, or the activation of signal transduction pathways that regulate plant defense in response to attacks by herbivorous pests (including, for example, protease inhibitors).

Increased tolerance of plants to nutrient-limited soils. Presently disclosed transcription factor genes introduced into plants may provide the means to improve uptake of essential nutrients, including nitrogenous compounds, phosphates, potassium, and trace minerals. The effect of these modifications is to increase the seedling germination and range of ornamental and crop plants. The utilities of presently disclosed transcription factor genes conferring tolerance to conditions of low nutrients also include cost savings to the grower by reducing the amounts of fertilizer needed, environmental benefits of reduced fertilizer runoff; and improved yield and stress tolerance. In addition, this gene could be used to alter seed protein amounts and/or composition that could impact yield as well as the nutritional value and production of various food products.

Hormone sensitivity. One or more of the presently disclosed transcription factor genes have been shown to affect plant abscisic acid (ABA) sensitivity. This plant hormone is likely the most important hormone in mediating the adaptation of a plant to stress. For example, ABA mediates conversion of apical meristems into dormant buds. In response to increasingly cold conditions, the newly developing leaves growing above the meristem become converted into stiff bud scales that closely wrap the meristem and protect it from mechanical damage during winter. ABA in the bud also enforces dormancy; during premature warm spells, the buds are inhibited from sprouting. Bud dormancy is eliminated after either a prolonged cold period of cold or a significant number of lengthening days. Thus, by affecting ABA sensitivity, introduced transcription factor genes may affect cold sensitivity and survivability. ABA is also important in protecting plants from drought tolerance.

Several other of the present transcription factor genes have been used to manipulate ethylene signal transduction and response pathways. These genes can thus be used to manipulate the processes influenced by ethylene, such as seed germination or fruit ripening, and to improve seed or fruit quality.

Production of seed and leaf prenyl lipids, including tocopherol. Prenyl lipids play a role in anchoring proteins in membranes or membranous organelles. Thus modifying the prenyl lipid content of seeds and leaves could affect membrane integrity and function. A number of presently disclosed transcription factor genes

have been shown to modify the tocopherol composition of plants. Tocopherols have both anti-oxidant and vitamin E activity.

Production of seed and leaf phytosterols: Presently disclosed transcription factor genes that modify levels of phytosterols in plants may have at least two utilities. First, phytosterols are an important source of precursors for the manufacture of human steroid hormones. Thus, regulation of transcription factor expression or activity could lead to elevated levels of important human steroid precursors for steroid semi-synthesis. For example, transcription factors that cause elevated levels of campesterol in leaves, or sitosterols and stigmasterols in seed crops, would be useful for this purpose. Phytosterols and their hydrogenated derivatives phytostanols also have proven cholesterol-lowering properties, and transcription factor genes that modify the expression of these compounds in plants would thus provide health benefits.

Production of seed and leaf glucosinolates. Some glucosinolates have anti-cancer activity; thus, increasing the levels or composition of these compounds by introducing several of the presently disclosed transcription factors might be of interest from a nutraceutical standpoint. (3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters might be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Modified seed oil content. The composition of seeds, particularly with respect to seed oil amounts and/or composition, is very important for the nutritional value and production of various food and feed products. Several of the presently disclosed transcription factor genes in seed lipid saturation that alter seed oil content could be used to improve the heat stability of oils or to improve the nutritional quality of seed oil, by, for example, reducing the number of calories in seed, increasing the number of calories in animal feeds, or altering the ratio of saturated to unsaturated lipids comprising the oils.

Seed and leaf fatty acid composition. A number of the presently disclosed transcription factor genes have been shown to alter the fatty acid composition in plants, and seeds in particular. This modification may find particular value for improving the nutritional value of, for example, seeds or whole plants. Dietary fatty acids ratios have been shown to have an effect on, for example, bone integrity and remodeling (see, for example, Weiler, H.A., *Pediatr Res* (2000) 47:5 692-697). The ratio of dietary fatty acids may alter the precursor pools of long-chain polyunsaturated fatty acids that serve as precursors for prostaglandin synthesis. In mammalian connective tissue, prostaglandins serve as important signals regulating the balance between resorption and formation in bone and cartilage. Thus dietary fatty acid ratios altered in seeds may affect the etiology and outcome of bone loss.

Modified seed protein content. As with seed oils, the composition of seeds, particularly with respect to protein amounts and/or composition, is very important for the nutritional value and production of various food and feed products. A number of the presently disclosed transcription factor genes modify the protein concentrations in seeds would provide nutritional benefits, and may be used to prolong storage, increase seed pest or disease resistance, or modify germination rates.

Production of flavonoids in leaves and other plant parts. Expression of presently disclosed transcription factor genes that increase flavonoid production in plants, including anthocyanins and condensed tannins, may be used to alter in pigment production for horticultural purposes, and possibly increasing stress resistance. Flavonoids have antimicrobial activity and could be used to engineer pathogen resistance. Several flavonoid compounds have health promoting effects such as the inhibition of tumor growth and cancer, prevention of bone loss and the prevention of the oxidation of lipids. Increasing levels of condensed tannins, whose biosynthetic pathway is shared with anthocyanin biosynthesis, in forage legumes is an important agronomic trait because they prevent pasture bloat by collapsing protein foams within the rumen. For a review on the utilities of flavonoids and their derivatives, refer to Dixon et al. (1999) *Trends Plant Sci.* 4:394-400.

Production of diterpenes in leaves and other plant parts. Depending on the plant species, varying amounts of diverse secondary biochemicals (often lipophilic

terpenes) are produced and exuded or volatilized by trichomes. These exotic secondary biochemicals, which are relatively easy to extract because they are on the surface of the leaf, have been widely used in such products as flavors and aromas, drugs, pesticides and cosmetics. Thus, the overexpression of genes that are used to produce diterpenes in plants may be accomplished by introducing transcription factor genes that induce said overexpression. One class of secondary metabolites, the diterpenes, can effect several biological systems such as tumor progression, prostaglandin synthesis and tissue inflammation. In addition, diterpenes can act as insect pheromones, termite allomones, and can exhibit neurotoxic, cytotoxic and antimitotic activities. As a result of this functional diversity, diterpenes have been the target of research several pharmaceutical ventures. In most cases where the metabolic pathways are impossible to engineer, increasing trichome density or size on leaves may be the only way to increase plant productivity.

Production of anthocyanin in leaves and other plant parts. Several presently disclosed transcription factor genes can be used to alter anthocyanin production in numerous plant species. The potential utilities of these genes include alterations in pigment production for horticultural purposes, and possibly increasing stress resistance in combination with another transcription factor.

Production of miscellaneous secondary metabolites. Microarray data suggests that flux through the aromatic amino acid biosynthetic pathways and primary and secondary metabolite biosynthetic pathways are up-regulated. Presently disclosed transcription factors have been shown to be involved in regulating alkaloid biosynthesis, in part by up-regulating the enzymes indole-3-glycerol phosphatase and strictosidine synthase. Phenylalanine ammonia lyase, chalcone synthase and trans-cinnamate mono-oxygenase are also induced, and are involved in phenylpropanoid biosynthesis.

Sugar, starch, hemicellulose composition. Overexpression of the presently disclosed transcription factors that affect sugar content resulted in plants with altered leaf insoluble sugar content. Transcription factors that alter plant cell wall composition have several potential applications including altering food digestibility, plant tensile strength, wood quality, pathogen resistance and in pulp production. The

potential utilities of a gene involved in glucose-specific sugar sensing are to alter energy balance, photosynthetic rate, carbohydrate accumulation, biomass production, source-sink relationships, and senescence.

Hemicellulose is not desirable in paper pulps because of its lack of strength compared with cellulose. Thus modulating the amounts of cellulose vs. hemicellulose in the plant cell wall is desirable for the paper/lumber industry. Increasing the insoluble carbohydrate content in various fruits, vegetables, and other edible consumer products will result in enhanced fiber content. Increased fiber content would not only provide health benefits in food products, but might also increase digestibility of forage crops. In addition, the hemicellulose and pectin content of fruits and berries affects the quality of jam and catsup made from them. Changes in hemicellulose and pectin content could result in a superior consumer product.

Plant response to sugars and sugar composition. In addition to their important role as an energy source and structural component of the plant cell, sugars are central regulatory molecules that control several aspects of plant physiology, metabolism and development. It is thought that this control is achieved by regulating gene expression and, in higher plants, sugars have been shown to repress or activate plant genes involved in many essential processes such as photosynthesis, glyoxylate metabolism, respiration, starch and sucrose synthesis and degradation, pathogen response, wounding response, cell cycle regulation, pigmentation, flowering and senescence. The mechanisms by which sugars control gene expression are not understood.

Because sugars are important signaling molecules, the ability to control either the concentration of a signaling sugar or how the plant perceives or responds to a signaling sugar could be used to control plant development, physiology or metabolism. For example, the flux of sucrose (a disaccharide sugar used for systemically transporting carbon and energy in most plants) has been shown to affect gene expression and alter storage compound accumulation in seeds. Manipulation of the sucrose signaling pathway in seeds may therefore cause seeds to have more protein, oil or carbohydrate, depending on the type of manipulation. Similarly, in tubers, sucrose is converted to starch which is used as an energy store. It is thought that sugar signaling pathways may partially determine the levels of starch synthesized

in the tubers. The manipulation of sugar signaling in tubers could lead to tubers with a higher starch content.

Thus, the presently disclosed transcription factor genes that manipulate the sugar signal transduction pathway may lead to altered gene expression to produce plants with desirable traits. In particular, manipulation of sugar signal transduction pathways could be used to alter source-sink relationships in seeds, tubers, roots and other storage organs leading to increase in yield.

Plant growth rate and development. A number of the presently disclosed transcription factor genes have been shown to have significant effects on plant growth rate and development. These observations have included, for example, more rapid or delayed growth and development of reproductive organs. This would provide utility for regions with short or long growing seasons, respectively. Accelerating plant growth would also improve early yield or increase biomass at an earlier stage, when such is desirable (for example, in producing forestry products).

Embryo development. Presently disclosed transcription factor genes that alter embryo development has been used to alter seed protein and oil amounts and/or composition which is very important for the nutritional value and production of various food products. Seed shape and seed coat may also be altered by these genes, which may provide for improved storage stability.

Seed germination rate. A number of the presently disclosed transcription factor genes have been shown to modify seed germination rate, including when the seeds are in conditions normally unfavorable for germination (e.g., cold, heat or salt stress, or in the presence of ABA), and may thus be used to modify and improve germination rates under adverse conditions.

Plant, seedling vigor. Seedlings transformed with presently disclosed transcription factors have been shown to possess larger cotyledons and appeared somewhat more advanced than control plants. This indicates that the seedlings developed more rapidly than the control plants. Rapid seedling development is likely to reduce loss due to diseases particularly prevalent at the seedling stage (e.g.,

damping off) and is thus important for survivability of plants germinating in the field or in controlled environments.

Senescence, cell death. Presently disclosed transcription factor genes may be used to alter senescence responses in plants. Although leaf senescence is thought to be an evolutionary adaptation to recycle nutrients, the ability to control senescence in an agricultural setting has significant value. For example, a delay in leaf senescence in some maize hybrids is associated with a significant increase in yields and a delay of a few days in the senescence of soybean plants can have a large impact on yield. Delayed flower senescence may also generate plants that retain their blossoms longer and this may be of potential interest to the ornamental horticulture industry.

Modified fertility. Plants that overexpress a number of the presently disclosed transcription factor genes have been shown to possess reduced fertility. This could be a desirable trait, as it could be exploited to prevent or minimize the escape of the pollen of genetically modified organisms (GMOs) into the environment.

Early and delayed flowering. Presently disclosed transcription factor genes that accelerate flowering could have valuable applications in such programs since they allow much faster generation times. In a number of species, for example, broccoli, cauliflower, where the reproductive parts of the plants constitute the crop and the vegetative tissues are discarded, it would be advantageous to accelerate time to flowering. Accelerating flowering could shorten crop and tree breeding programs. Additionally, in some instances, a faster generation time might allow additional harvests of a crop to be made within a given growing season. A number of *Arabidopsis* genes have already been shown to accelerate flowering when constitutively expressed. These include LEAFY, APETALA1 and CONSTANS (Mandel, M. et al., 1995, Nature 377, 522-524; Weigel, D. and Nilsson, O., 1995, Nature 377, 495-500; Simon et al., 1996, Nature 384, 59-62).

By regulating the expression of potential flowering using inducible promoters, flowering could be triggered by application of an inducer chemical. This would allow flowering to be synchronized across a crop and facilitate more efficient harvesting. Such inducible systems could also be used to tune the flowering of crop varieties to

different latitudes. At present, species such as soybean and cotton are available as a series of maturity groups that are suitable for different latitudes on the basis of their flowering time (which is governed by day-length). A system in which flowering could be chemically controlled would allow a single high-yielding northern maturity group to be grown at any latitude. In southern regions such plants could be grown for longer, thereby increasing yields, before flowering was induced. In more northern areas, the induction would be used to ensure that the crop flowers prior to the first winter frosts.

In a sizeable number of species, for example, root crops, where the vegetative parts of the plants constitute the crop and the reproductive tissues are discarded, it would be advantageous to delay or prevent flowering. Extending vegetative development with presently disclosed transcription factor genes could thus bring about large increases in yields.. Prevention of flowering might help maximize vegetative yields and prevent escape of genetically modified organism (GMO) pollen.

Extended flowering phase. Presently disclosed transcription factors that extend flowering time have utility in engineering plants with longer-lasting flowers for the horticulture industry, and for extending the time in which the plant is fertile.

Flower and leaf development. Presently disclosed transcription factor genes have been used to modify the development of flowers and leaves. This could be advantageous in the development of new ornamental cultivars that present unique configurations. In addition, some of these genes have been shown to reduce a plant's fertility, which is also useful for helping to prevent development of pollen of GMOs.

Flower abscission. Presently disclosed transcription factor genes introduced into plants have been used to retain flowers for longer periods. This would provide a significant benefit to the ornamental industry, for both cut flowers and woody plant varieties (of, for example, maize), as well as have the potential to lengthen the fertile period of a plant, which could positively impact yield and breeding programs.

A listing of specific effects and utilities that the presently disclosed transcription factor genes have on plants, as determined by direct observation and assay analysis, is provided in Table 4.

XVI. Antisense and Co-suppression

In addition to expression of the nucleic acids of the invention as gene replacement or plant phenotype modification nucleic acids, the nucleic acids are also useful for sense and anti-sense suppression of expression, e.g., to down-regulate expression of a nucleic acid of the invention, e.g., as a further mechanism for modulating plant phenotype. That is, the nucleic acids of the invention, or subsequences or anti-sense sequences thereof, can be used to block expression of naturally occurring homologous nucleic acids. A variety of sense and anti-sense technologies are known in the art, e.g., as set forth in Lichtenstein and Nellen (1997) Antisense Technology: A Practical Approach IRL Press at Oxford University Press, Oxford, U.K.. In general, sense or anti-sense sequences are introduced into a cell, where they are optionally amplified, e.g., by transcription. Such sequences include both simple oligonucleotide sequences and catalytic sequences such as ribozymes.

For example, a reduction or elimination of expression (i.e., a "knock-out") of a transcription factor or transcription factor homologue polypeptide in a transgenic plant, e.g., to modify a plant trait, can be obtained by introducing an antisense construct corresponding to the polypeptide of interest as a cDNA. For antisense suppression, the transcription factor or homologue cDNA is arranged in reverse orientation (with respect to the coding sequence) relative to the promoter sequence in the expression vector. The introduced sequence need not be the full length cDNA or gene, and need not be identical to the cDNA or gene found in the plant type to be transformed. Typically, the antisense sequence need only be capable of hybridizing to the target gene or RNA of interest. Thus, where the introduced sequence is of shorter length, a higher degree of homology to the endogenous transcription factor sequence will be needed for effective antisense suppression. While antisense sequences of various lengths can be utilized, preferably, the introduced antisense sequence in the vector will be at least 30 nucleotides in length, and improved antisense suppression will typically be observed as the length of the antisense sequence increases. Preferably, the length of the antisense sequence in the vector will be greater than 100 nucleotides.

Transcription of an antisense construct as described results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous transcription factor gene in the plant cell.

Suppression of endogenous transcription factor gene expression can also be achieved using a ribozyme. Ribozymes are RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 and U.S. Patent No. 5,543,508. Synthetic ribozyme sequences including antisense RNAs can be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that hybridize to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

Suppression of endogenous transcription factor gene expression can also be achieved using RNA interference, or RNAi. RNAi is a post-transcriptional, targeted gene-silencing technique that uses double-stranded RNA (dsRNA) to incite degradation of messenger RNA (mRNA) containing the same sequence as the dsRNA (Constans, (2002) *The Scientist* 16:36). Small interfering RNAs, or siRNAs are produced in at least two steps: an endogenous ribonuclease cleaves longer dsRNA into shorter, 21-23 nucleotide-long RNAs. The siRNA segments then mediate the degradation of the target mRNA (Zamore, (2001) *Nature Struct. Biol.*, 8:746-50). RNAi has been used for gene function determination in a manner similar to antisense oligonucleotides (Constans, (2002) *The Scientist* 16:36). Expression vectors that continually express siRNAs in transiently and stably transfected have been engineered to express small hairpin RNAs (shRNAs), which get processed in vivo into siRNA-like molecules capable of carrying out gene-specific silencing (Brummelkamp et al., (2002) *Science* 296:550-553, and Paddison, et al. (2002) *Genes & Dev.* 16:948-958). Post-transcriptional gene silencing by double-stranded RNA is discussed in further detail by Hammond et al. (2001) *Nature Rev Gen* 2: 110-119, Fire et al. (1998) *Nature* 391: 806-811 and Timmons and Fire (1998) *Nature* 395: 854.

Vectors in which RNA encoded by a transcription factor or transcription factor homologue cDNA is over-expressed can also be used to obtain co-suppression of a corresponding endogenous gene, e.g., in the manner described in U.S. Patent No.

5,231,020 to Jorgensen. Such co-suppression (also termed sense suppression) does not require that the entire transcription factor cDNA be introduced into the plant cells, nor does it require that the introduced sequence be exactly identical to the endogenous transcription factor gene of interest. However, as with antisense suppression, the suppressive efficiency will be enhanced as specificity of hybridization is increased, e.g., as the introduced sequence is lengthened, and/or as the sequence similarity between the introduced sequence and the endogenous transcription factor gene is increased.

Vectors expressing an untranslatable form of the transcription factor mRNA, e.g., sequences comprising one or more stop codon, or nonsense mutation) can also be used to suppress expression of an endogenous transcription factor, thereby reducing or eliminating its activity and modifying one or more traits. Methods for producing such constructs are described in U.S. Patent No. 5,583,021. Preferably, such constructs are made by introducing a premature stop codon into the transcription factor gene. Alternatively, a plant trait can be modified by gene silencing using double-strand RNA (Sharp (1999) Genes and Development 13: 139-141). Another method for abolishing the expression of a gene is by insertion mutagenesis using the T-DNA of *Agrobacterium tumefaciens*. After generating the insertion mutants, the mutants can be screened to identify those containing the insertion in a transcription factor or transcription factor homologue gene. Plants containing a single transgene insertion event at the desired gene can be crossed to generate homozygous plants for the mutation. Such methods are well known to those of skill in the art. (See for example Koncz et al. (1992) Methods in Arabidopsis Research, World Scientific.)

Alternatively, a plant phenotype can be altered by eliminating an endogenous gene, such as a transcription factor or transcription factor homologue, e.g., by homologous recombination (Kempin et al. (1997) Nature 389:802-803).

A plant trait can also be modified by using the Cre-lox system (for example, as described in US Pat. No. 5,658,772). A plant genome can be modified to include first and second lox sites that are then contacted with a Cre recombinase. If the lox sites are in the same orientation, the intervening DNA sequence between the two sites

is excised. If the lox sites are in the opposite-orientation, the intervening sequence is inverted.

The polynucleotides and polypeptides of this invention can also be expressed in a plant in the absence of an expression cassette by manipulating the activity or expression level of the endogenous gene by other means. For example, by ectopically expressing a gene by T-DNA activation tagging (Ichikawa et al. (1997) Nature 390 698-701; Kakimoto et al. (1996) Science 274: 982-985). This method entails transforming a plant with a gene tag containing multiple transcriptional enhancers and once the tag has inserted into the genome, expression of a flanking gene coding sequence becomes deregulated. In another example, the transcriptional machinery in a plant can be modified so as to increase transcription levels of a polynucleotide of the invention (*See, e.g.,* PCT Publications WO 96/06166 and WO 98/53057 which describe the modification of the DNA-binding specificity of zinc finger proteins by changing particular amino acids in the DNA-binding motif).

The transgenic plant can also include the machinery necessary for expressing or altering the activity of a polypeptide encoded by an endogenous gene, for example by altering the phosphorylation state of the polypeptide to maintain it in an activated state.

Transgenic plants (or plant cells, or plant explants, or plant tissues) incorporating the polynucleotides of the invention and/or expressing the polypeptides of the invention can be produced by a variety of well established techniques as described above. Following construction of a vector, most typically an expression cassette, including a polynucleotide, e.g., encoding a transcription factor or transcription factor homologue, of the invention, standard techniques can be used to introduce the polynucleotide into a plant, a plant cell, a plant explant or a plant tissue of interest. Optionally, the plant cell, explant or tissue can be regenerated to produce a transgenic plant.

The plant can be any higher plant, including gymnosperms, monocotyledonous and dicotyledonous plants. Suitable protocols are available for *Leguminosae* (alfalfa, soybean, clover, etc.), *Umbelliferae* (carrot, celery, parsnip),

Cruciferae (cabbage, radish, rapeseed, broccoli, etc.), *Curcubitaceae* (melons and cucumber), *Gramineae* (wheat, corn, rice, barley, millet, etc.), *Solanaceae* (potato, tomato, tobacco, peppers, etc.), and various other crops. See protocols described in Ammirato et al. (1984) Handbook of Plant Cell Culture –Crop Species, Macmillan Publ. Co. Shimamoto et al. (1989) Nature 338:274-276; Fromm et al. (1990) Bio/Technology 8:833-839; and Vasil et al. (1990) Bio/Technology 8:429-434.

Transformation and regeneration of both monocotyledonous and dicotyledonous plant cells is now routine, and the selection of the most appropriate transformation technique will be determined by the practitioner. The choice of method will vary with the type of plant to be transformed; those skilled in the art will recognize the suitability of particular methods for given plant types. Suitable methods can include, but are not limited to: electroporation of plant protoplasts; liposome-mediated transformation; polyethylene glycol (PEG) mediated transformation; transformation using viruses; micro-injection of plant cells; micro-projectile bombardment of plant cells; vacuum infiltration; and *Agrobacterium tumefaciens* mediated transformation. Transformation means introducing a nucleotide sequence into a plant in a manner to cause stable or transient expression of the sequence.

Successful examples of the modification of plant characteristics by transformation with cloned sequences which serve to illustrate the current knowledge in this field of technology, and which are herein incorporated by reference, include: U.S. Patent Nos. 5,571,706; 5,677,175; 5,510,471; 5,750,386; 5,597,945; 5,589,615; 5,750,871; 5,268,526; 5,780,708; 5,538,880; 5,773,269; 5,736,369 and 5,610,042.

Following transformation, plants are preferably selected using a dominant selectable marker incorporated into the transformation vector. Typically, such a marker will confer antibiotic or herbicide resistance on the transformed plants, and selection of transformants can be accomplished by exposing the plants to appropriate concentrations of the antibiotic or herbicide.

After transformed plants are selected and grown to maturity, those plants showing a modified trait are identified. The modified trait can be any of those traits described above. Additionally, to confirm that the modified trait is due to changes in

expression levels or activity of the polypeptide or polynucleotide of the invention can be determined by analyzing mRNA expression using Northern blots, RT-PCR or microarrays, or protein expression using immunoblots or Western blots or gel shift assays.

XVII. Integrated Systems – Sequence Identity

Additionally, the present invention may be an integrated system, computer or computer readable medium that comprises an instruction set for determining the identity of one or more sequences in a database. In addition, the instruction set can be used to generate or identify sequences that meet any specified criteria. Furthermore, the instruction set may be used to associate or link certain functional benefits, such improved characteristics, with one or more identified sequence.

For example, the instruction set can include, e.g., a sequence comparison or other alignment program, e.g., an available program such as, for example, the Wisconsin Package Version 10.0, such as BLAST, FASTA, PILEUP, FINDPATTERNS or the like (GCG, Madison, WI). Public sequence databases such as GenBank, EMBL, Swiss-Prot and PIR or private sequence databases such as PHYTOSEQ sequence database (Incyte Genomics, Palo Alto, CA) can be searched.

Alignment of sequences for comparison can be conducted by the local homology algorithm of Smith and Waterman (1981) Adv. Appl. Math. 2:482, by the homology alignment algorithm of Needleman and Wunsch (1970) J. Mol. Biol. 48:443-453, by the search for similarity method of Pearson and Lipman (1988) Proc. Natl. Acad. Sci. U.S.A. 85:2444-2448, by computerized implementations of these algorithms. After alignment, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two sequences over a comparison window to identify and compare local regions of sequence similarity. The comparison window can be a segment of at least about 20 contiguous positions, usually about 50 to about 200, more usually about 100 to about 150 contiguous positions. A description of the method is provided in Ausubel et al., *supra*.

A variety of methods for determining sequence relationships can be used, including manual alignment and computer assisted sequence alignment and analysis. This later approach is a preferred approach in the present invention, due to the increased throughput afforded by computer assisted methods. As noted above, a variety of computer programs for performing sequence alignment are available, or can be produced by one of skill.

One example algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul et al. J. Mol. Biol. 215:403-410 (1990). Software for performing BLAST analyses is publicly available, e.g., through the National Center for Biotechnology Information (see internet website at ncbi.nlm.nih.gov). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul et al., *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) Proc. Natl. Acad. Sci. USA 89:10915). Unless otherwise indicated, "sequence identity" here refers to the % sequence identity generated from a tblastx using the NCBI version of

the algorithm at the default settings using gapped alignments with the filter "off" (see, for example, internet website at ncbi.nlm.nih.gov).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (*see, e.g.,* Karlin & Altschul (1993) Proc. Natl. Acad. Sci. USA 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability ($P(N)$), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence (and, therefore, in this context, homologous) if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, or less than about 0.01, and or even less than about 0.001. An additional example of a useful sequence alignment algorithm is PILEUP. PILEUP creates a multiple sequence alignment from a group of related sequences using progressive, pairwise alignments. The program can align, e.g., up to 300 sequences of a maximum length of 5,000 letters.

The integrated system, or computer typically includes a user input interface allowing a user to selectively view one or more sequence records corresponding to the one or more character strings, as well as an instruction set which aligns the one or more character strings with each other or with an additional character string to identify one or more region of sequence similarity. The system may include a link of one or more character strings with a particular phenotype or gene function. Typically, the system includes a user readable output element that displays an alignment produced by the alignment instruction set.

The methods of this invention can be implemented in a localized or distributed computing environment. In a distributed environment, the methods may implemented on a single computer comprising multiple processors or on a multiplicity of computers. The computers can be linked, e.g. through a common bus, but more preferably the computer(s) are nodes on a network. The network can be a generalized or a dedicated local or wide-area network and, in certain preferred embodiments, the computers may be components of an intra-net or an internet.

Thus, the invention provides methods for identifying a sequence similar or homologous to one or more polynucleotides as noted herein, or one or more target polypeptides encoded by the polynucleotides, or otherwise noted herein and may include linking or associating a given plant phenotype or gene function with a sequence. In the methods, a sequence database is provided (locally or across an inter or intra net) and a query is made against the sequence database using the relevant sequences herein and associated plant phenotypes or gene functions.

Any sequence herein can be entered into the database, before or after querying the database. This provides for both expansion of the database and, if done before the querying step, for insertion of control sequences into the database. The control sequences can be detected by the query to ensure the general integrity of both the database and the query. As noted, the query can be performed using a web browser based interface. For example, the database can be a centralized public database such as those noted herein, and the querying can be done from a remote terminal or computer across an internet or intranet.

XVIII. Examples

The following examples are intended to illustrate but not limit the present invention. The complete descriptions of the traits associated with each polynucleotide of the invention is fully disclosed in Table 4 and Table 6.

Example I: Full Length Gene Identification and Cloning

Putative transcription factor sequences (genomic or ESTs) related to known transcription factors were identified in the *Arabidopsis thaliana* GenBank database using the tblastn sequence analysis program using default parameters and a P-value cutoff threshold of -4 or -5 or lower, depending on the length of the query sequence. Putative transcription factor sequence hits were then screened to identify those containing particular sequence strings. If the sequence hits contained such sequence strings, the sequences were confirmed as transcription factors.

Alternatively, *Arabidopsis thaliana* cDNA libraries derived from different tissues or treatments, or genomic libraries were screened to identify novel members of a transcription family using a low stringency hybridization approach. Probes were

synthesized using gene specific primers in a standard PCR reaction (annealing temperature 60° C) and labeled with ³²P dCTP using the High Prime DNA Labeling Kit (Boehringer Mannheim). Purified radiolabelled probes were added to filters immersed in Church hybridization medium (0.5 M NaPO₄ pH 7.0, 7% SDS, 1 % w/v bovine serum albumin) and hybridized overnight at 60°C with shaking. Filters were washed two times for 45 to 60 minutes with 1xSCC, 1% SDS at 60° C.

To identify additional sequence 5' or 3' of a partial cDNA sequence in a cDNA library, 5' and 3' rapid amplification of cDNA ends (RACE) was performed using the U.C. Marathon cDNA amplification kit (Clontech, Palo Alto, CA). Generally, the method entailed first isolating poly(A) mRNA, performing first and second strand cDNA synthesis to generate double stranded cDNA, blunting cDNA ends, followed by ligation of the U.C. Marathon Adaptor to the cDNA to form a library of adaptor-ligated ds cDNA.

Gene-specific primers were designed to be used along with adaptor specific primers for both 5' and 3' RACE reactions. Nested primers, rather than single primers, were used to increase PCR specificity. Using 5' and 3' RACE reactions, 5' and 3' RACE fragments were obtained, sequenced and cloned. The process can be repeated until 5' and 3' ends of the full-length gene were identified. Then the full-length cDNA was generated by PCR using primers specific to 5' and 3' ends of the gene by end-to-end PCR.

Example II: Construction of Expression Vectors

The sequence was amplified from a genomic or cDNA library using primers specific to sequences upstream and downstream of the coding region. The expression vector was pMEN20 or pMEN65, which are both derived from pMON316 (Sanders et al, (1987) Nucleic Acids Research 15:1543-1558) and contain the CaMV 35S promoter to express transgenes. To clone the sequence into the vector, both pMEN20 and the amplified DNA fragment were digested separately with SalI and NotI restriction enzymes at 37° C for 2 hours. The digestion products were subject to electrophoresis in a 0.8% agarose gel and visualized by ethidium bromide staining. The DNA fragments containing the sequence and the linearized plasmid were excised and purified by using a Qiaquick gel extraction kit (Qiagen, Valencia CA). The

fragments of interest were ligated at a ratio of 3:1 (vector to insert). Ligation reactions using T4 DNA ligase (New England Biolabs, Beverly MA) were carried out at 16° C for 16 hours. The ligated DNAs were transformed into competent cells of the *E. coli* strain DH5alpha by using the heat shock method. The transformations were plated on LB plates containing 50 mg/l kanamycin (Sigma, St. Louis, MO). Individual colonies were grown overnight in five milliliters of LB broth containing 50 mg/l kanamycin at 37° C. Plasmid DNA was purified by using Qiaquick Mini Prep kits (Qiagen).

Example III: Transformation of *Agrobacterium* with the Expression Vector

After the plasmid vector containing the gene was constructed, the vector was used to transform *Agrobacterium tumefaciens* cells expressing the gene products. The stock of *Agrobacterium tumefaciens* cells for transformation were made as described by Nagel et al. (1990) FEMS Microbiol Letts. 67: 325-328. *Agrobacterium* strain ABI was grown in 250 ml LB medium (Sigma) overnight at 28°C with shaking until an absorbance (A_{600}) of 0.5 – 1.0 was reached. Cells were harvested by centrifugation at 4,000 x g for 15 min at 4° C. Cells were then resuspended in 250 µl chilled buffer (1 mM HEPES, pH adjusted to 7.0 with KOH). Cells were centrifuged again as described above and resuspended in 125 µl chilled buffer. Cells were then centrifuged and resuspended two more times in the same HEPES buffer as described above at a volume of 100 µl and 750 µl, respectively. Resuspended cells were then distributed into 40 µl aliquots, quickly frozen in liquid nitrogen, and stored at -80° C.

Agrobacterium cells were transformed with plasmids prepared as described above following the protocol described by Nagel et al. For each DNA construct to be transformed, 50 – 100 ng DNA (generally resuspended in 10 mM Tris-HCl, 1 mM EDTA, pH 8.0) was mixed with 40 µl of *Agrobacterium* cells. The DNA/cell mixture was then transferred to a chilled cuvette with a 2mm electrode gap and subject to a 2.5 kV charge dissipated at 25 µF and 200 µF using a Gene Pulser II apparatus (Bio-Rad, Hercules, CA). After electroporation, cells were immediately resuspended in 1.0 ml LB and allowed to recover without antibiotic selection for 2 – 4 hours at 28° C in a shaking incubator. After recovery, cells were plated onto selective medium of LB broth containing 100 µg/ml spectinomycin (Sigma) and incubated for 24-48 hours at 28° C. Single colonies were then picked and inoculated in fresh medium. The

presence of the plasmid construct was verified by PCR amplification and sequence analysis.

Example IV: Transformation of *Arabidopsis* Plants with *Agrobacterium tumefaciens* with Expression Vector

After transformation of *Agrobacterium tumefaciens* with plasmid vectors containing the gene, single *Agrobacterium* colonies were identified, propagated, and used to transform *Arabidopsis* plants. Briefly, 500 ml cultures of LB medium containing 50 mg/l kanamycin were inoculated with the colonies and grown at 28° C with shaking for 2 days until an optical absorbance at 600 nm wavelength over 1 cm (A_{600}) of > 2.0 is reached. Cells were then harvested by centrifugation at 4,000 x g for 10 min, and resuspended in infiltration medium (1/2 X Murashige and Skoog salts (Sigma), 1 X Gamborg's B-5 vitamins (Sigma), 5.0% (w/v) sucrose (Sigma), 0.044 μ M benzylamino purine (Sigma), 200 μ l/l Silwet L-77 (Lehle Seeds) until an A_{600} of 0.8 was reached.

Prior to transformation, *Arabidopsis thaliana* seeds (ecotype Columbia) were sown at a density of ~10 plants per 4" pot onto Pro-Mix BX potting medium (Hummert International) covered with fiberglass mesh (18 mm X 16 mm). Plants were grown under continuous illumination (50-75 μ E/m²/sec) at 22-23° C with 65-70% relative humidity. After about 4 weeks, primary inflorescence stems (bolts) are cut off to encourage growth of multiple secondary bolts. After flowering of the mature secondary bolts, plants were prepared for transformation by removal of all siliques and opened flowers.

The pots were then immersed upside down in the mixture of *Agrobacterium* infiltration medium as described above for 30 sec, and placed on their sides to allow draining into a 1' x 2' flat surface covered with plastic wrap. After 24 h, the plastic wrap was removed and pots are turned upright. The immersion procedure was repeated one week later, for a total of two immersions per pot. Seeds were then collected from each transformation pot and analyzed following the protocol described below.

Example V: Identification of *Arabidopsis* Primary Transformants

Seeds collected from the transformation pots were sterilized essentially as follows. Seeds were dispersed into in a solution containing 0.1% (v/v) Triton X-100 (Sigma) and sterile H₂O and washed by shaking the suspension for 20 min. The wash solution was then drained and replaced with fresh wash solution to wash the seeds for 20 min with shaking. After removal of the second wash solution, a solution containing 0.1% (v/v) Triton X-100 and 70% ethanol (Equistar) was added to the seeds and the suspension was shaken for 5 min. After removal of the ethanol/detergent solution, a solution containing 0.1% (v/v) Triton X-100 and 30% (v/v) bleach (Clorox) was added to the seeds, and the suspension was shaken for 10 min. After removal of the bleach/detergent solution, seeds were then washed five times in sterile distilled H₂O. The seeds were stored in the last wash water at 4° C for 2 days in the dark before being plated onto antibiotic selection medium (1 X Murashige and Skoog salts (pH adjusted to 5.7 with 1M KOH), 1 X Gamborg's B-5 vitamins, 0.9% phytagar (Life Technologies), and 50 mg/l kanamycin). Seeds were germinated under continuous illumination (50-75 $\mu\text{E}/\text{m}^2/\text{sec}$) at 22-23° C. After 7-10 days of growth under these conditions, kanamycin resistant primary transformants (T₁ generation) were visible and obtained. These seedlings were transferred first to fresh selection plates where the seedlings continued to grow for 3-5 more days, and then to soil (Pro-Mix BX potting medium).

Primary transformants were crossed and progeny seeds (T₂) collected; kanamycin resistant seedlings were selected and analyzed. The expression levels of the recombinant polynucleotides in the transformants varies from about a 5% expression level increase to a least a 100% expression level increase. Similar observations are made with respect to polypeptide level expression.

Example VI: Identification of *Arabidopsis* Plants with Transcription Factor Gene Knockouts

The screening of insertion mutagenized *Arabidopsis* collections for null mutants in a known target gene was essentially as described in Krysan et al (1999) Plant Cell 11:2283-2290. Briefly, gene-specific primers, nested by 5-250 base pairs to each other, were designed from the 5' and 3' regions of a known target gene. Similarly, nested sets of primers were also created specific to each of the T-DNA or

transposon ends (the "right" and "left" borders). All possible combinations of gene specific and T-DNA/transposon primers were used to detect by PCR an insertion event within or close to the target gene. The amplified DNA fragments were then sequenced which allows the precise determination of the T-DNA/transposon insertion point relative to the target gene. Insertion events within the coding or intervening sequence of the genes were deconvoluted from a pool comprising a plurality of insertion events to a single unique mutant plant for functional characterization. The method is described in more detail in Yu and Adam, US Application Serial No. 09/177,733 filed October 23, 1998.

Example VII: Identification of Modified Phenotypes in Overexpression or Gene Knockout Plants

Experiments were performed to identify those transformants or knockouts that exhibited modified biochemical characteristics. Among the biochemicals that were assayed were insoluble sugars, such as arabinose, fucose, galactose, mannose, rhamnose or xylose or the like; prenyl lipids, such as lutein, beta-carotene, xanthophyll-1, xanthophyll-2, chlorophylls A or B, or alpha-, delta- or gamma-tocopherol or the like; fatty acids, such as 16:0 (palmitic acid), 16:1 (palmitoleic acid), 18:0 (stearic acid), 18:1 (oleic acid), 18:2 (linoleic acid), 20:0, 18:3 (linolenic acid), 20:1 (eicosenoic acid), 20:2, 22:1 (erucic acid) or the like; waxes, such as by altering the levels of C29, C31, or C33 alkanes; sterols, such as brassicasterol, campesterol, stigmasterol, sitosterol or stigmastanol or the like, glucosinolates, protein or oil levels.

Fatty acids were measured using two methods depending on whether the tissue was from leaves or seeds. For leaves, lipids were extracted and esterified with hot methanolic H_2SO_4 and partitioned into hexane from methanolic brine. For seed fatty acids, seeds were pulverized and extracted in methanol:heptane:toluene:2,2-dimethoxypropane: H_2SO_4 (39:34:20:5:2) for 90 minutes at 80°C. After cooling to room temperature the upper phase, containing the seed fatty acid esters, was subjected to GC analysis. Fatty acid esters from both seed and leaf tissues were analyzed with a Supelco SP-2330 column.

Glucosinolates were purified from seeds or leaves by first heating the tissue at 95°C for 10 minutes. Preheated ethanol:water (50:50) is and after heating at 95°C for a further 10 minutes, the extraction solvent is applied to a DEAE Sephadex column which had been previously equilibrated with 0.5 M pyridine acetate. Desulfoglucosinolates were eluted with 300 ul water and analyzed by reverse phase HPLC monitoring at 226 nm.

For wax alkanes, samples were extracted using an identical method as fatty acids and extracts were analyzed on a HP 5890 GC coupled with a 5973 MSD. Samples were chromatographically isolated on a J&W DB35 mass spectrometer (J&W Scientific).

To measure prenyl lipids levels, seeds or leaves were pulverized with 1 to 2% pyrogallol as an antioxidant. For seeds, extracted samples were filtered and a portion removed for tocopherol and carotenoid/chlorophyll analysis by HPLC. The remaining material was saponified for sterol determination. For leaves, an aliquot was removed and diluted with methanol and chlorophyll A, chlorophyll B, and total carotenoids measured by spectrophotometry by determining optical absorbance at 665.2 nm, 652.5 nm, and 470 nm. An aliquot was removed for tocopherol and carotenoid/chlorophyll composition by HPLC using a Waters uBondapak C18 column (4.6 mm x 150 mm). The remaining methanolic solution was saponified with 10% KOH at 80°C for one hour. The samples were cooled and diluted with a mixture of methanol and water. A solution of 2% methylene chloride in hexane was mixed in and the samples were centrifuged. The aqueous methanol phase was again re-extracted 2% methylene chloride in hexane and, after centrifugation, the two upper phases were combined and evaporated. 2% methylene chloride in hexane was added to the tubes and the samples were then extracted with one ml of water. The upper phase was removed, dried, and resuspended in 400 ul of 2% methylene chloride in hexane and analyzed by gas chromatography using a 50 m DB-5ms (0.25 mm ID, 0.25 um phase, J&W Scientific).

Insoluble sugar levels were measured by the method essentially described by Reiter et al., (1997) Plant Journal 12:335-345. This method analyzes the neutral sugar composition of cell wall polymers found in *Arabidopsis* leaves. Soluble sugars were

separated from sugar polymers by extracting leaves with hot 70% ethanol. The remaining residue containing the insoluble polysaccharides was then acid hydrolyzed with allose added as an internal standard. Sugar monomers generated by the hydrolysis were then reduced to the corresponding alditols by treatment with NaBH₄, then were acetylated to generate the volatile alditol acetates which were then analyzed by GC-FID. Identity of the peaks was determined by comparing the retention times of known sugars converted to the corresponding alditol acetates with the retention times of peaks from wild-type plant extracts. Alditol acetates were analyzed on a Supelco SP-2330 capillary column (30 m x 250 μ m x 0.2 μ m) using a temperature program beginning at 180° C for 2 minutes followed by an increase to 220° C in 4 minutes. After holding at 220° C for 10 minutes, the oven temperature is increased to 240° C in 2 minutes and held at this temperature for 10 minutes and brought back to room temperature.

To identify plants with alterations in total seed oil or protein content, 150mg of seeds from T2 progeny plants were subjected to analysis by Near Infrared Reflectance Spectroscopy (NIRS) using a Foss NirSystems Model 6500 with a spinning cup transport system. NIRS is a non-destructive analytical method used to determine seed oil and protein composition. Infrared is the region of the electromagnetic spectrum located after the visible region in the direction of longer wavelengths. 'Near infrared' owns its name for being the infrared region near to the visible region of the electromagnetic spectrum. For practical purposes, near infrared comprises wavelengths between 800 and 2500 nm. NIRS is applied to organic compounds rich in O-H bonds (such as moisture, carbohydrates, and fats), C-H bonds (such as organic compounds and petroleum derivatives), and N-H bonds (such as proteins and amino acids). The NIRS analytical instruments operate by statistically correlating NIRS signals at several wavelengths with the characteristic or property intended to be measured. All biological substances contain thousands of C-H, O-H, and N-H bonds. Therefore, the exposure to near infrared radiation of a biological sample, such as a seed, results in a complex spectrum which contains qualitative and quantitative information about the physical and chemical composition of that sample.

The numerical value of a specific analyte in the sample, such as protein content or oil content, is mediated by a calibration approach known as chemometrics.

Chemometrics applies statistical methods such as multiple linear regression (MLR), partial least squares (PLS), and principle component analysis (PCA) to the spectral data and correlates them with a physical property or other factor, that property or factor is directly determined rather than the analyte concentration itself. The method first provides "wet chemistry" data of the samples required to develop the calibration.

Calibration for Arabidopsis seed oil composition was performed using accelerated solvent extraction using 1 g seed sample size and was validated against certified canola seed. A similar wet chemistry approach was performed for seed protein composition calibration.

Data obtained from NIRS analysis was analyzed statistically using a nearest-neighbor (N-N) analysis. The N-N analysis allows removal of within-block spatial variability in a fairly flexible fashion which does not require prior knowledge of the pattern of variability in the chamber. Ideally, all hybrids are grown under identical experimental conditions within a block (rep). In reality, even in many block designs, significant within-block variability exists. Nearest-neighbor procedures are based on assumption that environmental effect of a plot is closely related to that of its neighbors. Nearest-neighbor methods use information from adjacent plots to adjust for within-block heterogeneity and so provide more precise estimates of treatment means and differences. If there is within-plot heterogeneity on a spatial scale that is larger than a single plot and smaller than the entire block, then yields from adjacent plots will be positively correlated. Information from neighboring plots can be used to reduce or remove the unwanted effect of the spatial heterogeneity, and hence improve the estimate of the treatment effect. Data from neighboring plots can also be used to reduce the influence of competition between adjacent plots. The Papadakis N-N analysis can be used with designs to remove within-block variability that would not be removed with the standard split plot analysis (Papadakis, 1973, Inst. d'Amelior. Plantes Thessaloniki (Greece) Bull. Scientif., No. 23; Papadakis, 1984, Proc. Acad. Athens, 59, 326-342).

Experiments were performed to identify those transformants or knockouts that exhibited an improved pathogen tolerance. For such studies, the transformants were exposed to biotrophic fungal pathogens, such as *Erysiphe orontii*, and necrotropic

fungal pathogens, such as *Fusarium oxysporum*. *Fusarium oxysporum* isolates cause vascular wilts and damping off of various annual vegetables, perennials and weeds (Mauch-Mani and Slusarenko (1994) Molecular Plant-Microbe Interactions 7: 378-383). For *Fusarium oxysporum* experiments, plants grown on Petri dishes were sprayed with a fresh spore suspension of *F. oxysporum*. The spore suspension was prepared as follows: A plug of fungal hyphae from a plate culture was placed on a fresh potato dextrose agar plate and allowed to spread for one week. 5 ml sterile water was then added to the plate, swirled, and pipetted into 50 ml Armstrong *Fusarium* medium. Spores were grown overnight in *Fusarium* medium and then sprayed onto plants using a Preval paint sprayer. Plant tissue was harvested and frozen in liquid nitrogen 48 hours post infection.

Erysiphe orontii is a causal agent of powdery mildew. For *Erysiphe orontii* experiments, plants were grown approximately 4 weeks in a greenhouse under 12 hour light (20°C, ~30% relative humidity (rh)). Individual leaves were infected with *E. orontii* spores from infected plants using a camel's hair brush, and the plants were transferred to a Percival growth chamber (20°C, 80% rh.). Plant tissue was harvested and frozen in liquid nitrogen 7 days post infection.

Botrytis cinerea is a necrotrophic pathogen. *Botrytis cinerea* was grown on potato dextrose agar in the light. A spore culture was made by spreading 10 ml of sterile water on the fungus plate, swirling and transferring spores to 10 ml of sterile water. The spore inoculum (approx. 105 spores/ml) was used to spray 10 day-old seedlings grown under sterile conditions on MS (minus sucrose) media. Symptoms were evaluated every day up to approximately 1 week.

Infection with bacterial pathogens *Pseudomonas syringae* pv *maculicola* (Psm) strain 4326 and pv *maculicola* strain 4326 was performed by hand inoculation at two doses. Two inoculation doses allows the differentiation between plants with enhanced susceptibility and plants with enhanced resistance to the pathogen. Plants were grown for 3 weeks in the greenhouse, then transferred to the growth chamber for the remainder of their growth. Psm ES4326 was hand inoculated with 1 ml syringe on 3 fully-expanded leaves per plant (4 1/2 wk old), using at least 9 plants per overexpressing line at two inoculation doses, OD=0.005 and OD=0.0005. Disease

scoring occurred at day 3 post-inoculation with pictures of the plants and leaves taken in parallel.

In some instances, expression patterns of the pathogen-induced genes (such as defense genes) was monitored by microarray experiments. cDNAs were generated by PCR and resuspended at a final concentration of ~ 100 ng/ul in 3X SSC or 150mM Na-phosphate (Eisen and Brown (1999) *Methods Enzymol.* 303:179-205). The cDNAs were spotted on microscope glass slides coated with polylysine. The prepared cDNAs were aliquoted into 384 well plates and spotted on the slides using an x-y-z gantry (OmniGrid) purchased from GeneMachines (Menlo Park, CA) outfitted with quill type pins purchased from Telechem International (Sunnyvale, CA). After spotting, the arrays were cured for a minimum of one week at room temperature, rehydrated and blocked following the protocol recommended by Eisen and Brown (1999; *supra*).

Sample total RNA (10 ug) samples were labeled using fluorescent Cy3 and Cy5 dyes. Labeled samples were resuspended in 4X SSC/0.03% SDS/4 ug salmon sperm DNA/2 ug tRNA/ 50mM Na-pyrophosphate, heated for 95°C for 2.5 minutes, spun down and placed on the array. The array was then covered with a glass coverslip and placed in a sealed chamber. The chamber was then kept in a water bath at 62°C overnight. The arrays were washed as described in Eisen and Brown (1999) and scanned on a General Scanning 3000 laser scanner. The resulting files are subsequently quantified using Imogene, a software purchased from BioDiscovery (Los Angeles, CA).

Experiments were performed to identify those transformants or knockouts that exhibited an improved environmental stress tolerance. For such studies, the transformants were exposed to a variety of environmental stresses. Plants were exposed to chilling stress (6 hour exposure to 4-8°C), heat stress (6 hour exposure to 32-37°C), high salt stress (6 hour exposure to 200 mM NaCl), drought stress (168 hours after removing water from trays), osmotic stress (6 hour exposure to 3 M mannitol), or nutrient limitation (nitrogen, phosphate, and potassium) (Nitrogen: all components of MS medium remained constant except N was reduced to 20 mg/l of NH₄NO₃, or Phosphate: All components of MS medium except KH₂PO₄, which was

replaced by K_2SO_4 , Potassium: All components of MS medium except removal of KNO_3 and KH_2PO_4 , which were replaced by NaH_4PO_4).

Experiments were performed to identify those transformants or knockouts that exhibited a modified structure and development characteristics. For such studies, the transformants were observed by eye to identify novel structural or developmental characteristics associated with the ectopic expression of the polynucleotides or polypeptides of the invention.

Experiments were performed to identify those transformants or knockouts that exhibited modified sugar-sensing. For such studies, seeds from transformants were germinated on media containing 5% glucose or 9.4% sucrose which normally partially restrict hypocotyl elongation. Plants with altered sugar sensing may have either longer or shorter hypocotyls than normal plants when grown on this media. Additionally, other plant traits may be varied such as root mass.

Flowering time was measured by the number of rosette leaves present when a visible inflorescence of approximately 3 cm is apparent. Rosette and total leaf number on the progeny stem are tightly correlated with the timing of flowering (Koornneef et al (1991) *Mol. Gen. Genet* 229:57-66). The vernalization response was measured. For vernalization treatments, seeds were sown to MS agar plates, sealed with micropore tape, and placed in a 4°C cold room with low light levels for 6-8 weeks. The plates were then transferred to the growth rooms alongside plates containing freshly sown non-vernalized controls. Rosette leaves were counted when a visible inflorescence of approximately 3 cm was apparent.

Modified phenotypes observed for particular overexpressor or knockout plants are provided in Table 4. For a particular overexpressor that shows a less beneficial characteristic, it may be more useful to select a plant with a decreased expression of the particular transcription factor. For a particular knockout that shows a less beneficial characteristic, it may be more useful to select a plant with an increased expression of the particular transcription factor.

The sequences of the Sequence Listing or those in Tables 4, 5 or those disclosed here can be used to prepare transgenic plants and plants with altered traits. The specific transgenic plants listed below are produced from the sequences of the Sequence Listing, as noted. Table 4 provides exemplary polynucleotide and polypeptide sequences of the invention. Table 4 includes, from left to right for each sequence: the first column shows the polynucleotide SEQ ID NO; the second column shows the Mendel Gene ID No., GID; the third column shows the trait(s) resulting from the knock out or overexpression of the polynucleotide in the transgenic plant; the fourth column shows the category of the trait; the fifth column shows the transcription factor family to which the polynucleotide belongs; the sixth column ("Comment"), includes specific effects and utilities conferred by the polynucleotide of the first column; the seventh column shows the SEQ ID NO of the polypeptide encoded by the polynucleotide; and the eighth column shows the amino acid residue positions of the conserved domain in amino acid (AA) co-ordinates.

Seed of plants overexpressing sequences G265 (SEQ ID NOs:871 and 872), G715 (SEQ ID NOs:925 and 926), G1471 (SEQ ID NOs:311 and 312), G1793 (SEQ ID NOs:365 and 366), G1838 (SEQ ID NOs:381 and 382), G1902 (SEQ ID NOs:405 and 406), G286 (SEQ ID NOs:877 and 878), G2138 (SEQ ID NOs:865 and 866) and G2830 (SEQ ID NOs:875 and 876) was subjected to NIR analysis and a significant increase in seed oil content compared with seed from control plants was identified.

G192: G192 (SEQ ID NO: 859) was expressed in all plant tissues and under all conditions examined. Its expression was slightly induced upon infection by *Fusarium*. G192 was analyzed using transgenic plants in which this gene was expressed under the control of the 35S promoter. G192 overexpressors were late flowering under 12 hour light and had more leaves than control plants. This phenotype was manifested in the three T2 lines analyzed. Results of one experiment suggest that G192 overexpressor was more susceptible to infection with a moderate dose of the fungal pathogen *Erysiphe orontii*. The decrease in seed oil observed for one line was replicated in an independent experiment. G192 overexpression delayed flowering. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering, or for systems of inducible flowering time control. In particular, in species where the vegetative parts of the plants constitute the crop and

the reproductive tissues are discarded, it will be advantageous to delay or prevent flowering. Extending vegetative development can bring about large increases in yields. G192 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G192 can be used to manipulate seed oil content, which can be of nutritional value.

Closely Related Genes from Other Species

G192 had some similarity within the conserved WRKY domain to non-Arabidopsis plant proteins.

G1946: G1946 (SEQ ID NO: 801) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G1946 resulted in accelerated flowering, with 35S::G1946 transformants producing flower buds up to a week earlier than wild-type controls (24-hour light conditions). These effects were seen in 12/20 primary transformants and in two independent plantings of each of the three T2 lines. Unlike many early flowering Arabidopsis transgenic lines, which are dwarfed, 35S::G1946 transformants often reached full-size at maturity, and produced large quantities of seeds, although the plants were slightly pale in coloration and had slightly flat leaves compared to wild-type. In addition, 35S::G1946 plants showed an altered response to phosphate deprivation. Seedlings of G1946 overexpressor plants showed more secondary root growth on phosphate-free media, when compared to wild-type control. In a repeat experiment, all three lines showed the phenotype. Overexpression of G1946 in Arabidopsis also resulted in an increase in seed glucosinolate M39501 in T2 lines 1 and 3. An increase in seed oil and a decrease in seed protein was also observed in these two lines. G1946 was ubiquitously expressed, and does not appear to be significantly induced or repressed by any of the biotic and abiotic stress conditions tested at this time, with the exception of cold, which repressed G1946 expression. G1946 can be used to modify flowering time, as well as to improve the plant's performance in conditions of limited phosphate, and to alter seed oil, protein, and glucosinolate composition.

Closely Related Genes from Other Species

A comparison of the amino acid sequence of G1946 with sequences available from GenBank showed strong similarity with plant HSFs of several species (*Lycopersicon peruvianum*, *Medicago truncatula*, *Lycopersicon esculentum*, *Glycine max*, *Solanum tuberosum*, *Oryza sativa* and *Hordeum vulgare* subsp. *vulgare*).

G375: The sequence of G375 (SEQ ID NO:239) was experimentally determined and G375 was analyzed using transgenic plants in which G375 was expressed under the control of the 35S promoter. Overexpression of G375 produced marked effects on leaf development. At early stages of growth, 35S::G375 seedlings developed narrow, upward pointing leaves with long petioles (possibly indicating a disruption in circadian-clock controlled processes or nyctinastic movements). Additionally, some seedlings were noted to have elongated hypocotyls, and some were rather small compared to wild-type controls. Comparable phenotypes were obtained by overexpression of an AP2 family gene, G2113 (SEQ ID NO: 85). Following the switch to flowering, 35S::G375 plants showed reduced fertility, which possibly arose from a failure of stamens to fully elongate. One of the three T2 lines, (#41) was later flowering than wild-type controls, and also developed large numbers of small secondary rosette leaves in the axils of the primary rosette. Although these effects were not noted in the other two lines, the phenotypes obtained in line 41 were somewhat similar to those produced by overexpression of another Z-dof gene, G736 (SEQ ID NO: 211). G375 was expressed in all tissues, although at different levels. It was expressed at low levels in the root and germinating seed, and expressed at high levels in the embryo. The effects of G375 on leaf architecture are of potential interest to the ornamental horticulture industry.

Closely Related Genes from Other Species

G375 showed some homology to non-Arabidopsis plant proteins within the conserved Dof domain.

G1255: The sequence of G1255 (SEQ ID NO: 273) was experimentally determined and G1255 was analyzed using transgenic plants in which G1255 was expressed under the control of the 35S promoter. Plants overexpressing G1255 had

alterations in leaf architecture, a reduction in apical dominance, an increase in seed size, and showed more disease symptoms following inoculation with a low dose of the fungal pathogen *Botrytis cinerea*. G1255 was constitutively expressed and not significantly induced by any conditions tested. On the basis of the phenotypes produced by overexpression of G1255, G1255 can be used to manipulate the plant's defense response to produce pathogen resistance, alter plant architecture, or alter seed size.

Closely Related Genes from Other Species

G1255 showed strong homology to a putative rice zing finger protein represented by sequence AC087181_3. Sequence identity between these two protein extended beyond the conserved domain, and therefore, these genes can be orthologs.

G865: The complete cDNA sequence of G865 (SEQ ID NO: 557) was determined. G865 was ubiquitously expressed in Arabidopsis tissues. G865 was analyzed using transgenic plants in which G865 was expressed under the control of the 35S promoter. Plants overexpressing G865 were early flowering, with numerous secondary inflorescence meristems giving them a bushy appearance. G865 overexpressors were more susceptible to infection with a moderate dose of the fungal pathogens *Erysiphe orontii* and *Botrytis cinerea*. In addition, seeds from G865 overexpressing plants showed a trend of increased protein and reduced oil content, although the observed changes were not beyond the criteria used for judging significance except in one line. G865 can be used to control flowering time. G865 can be used to manipulate the defense response in order to generate pathogen-resistant plants. G865 can be used to alter seed oil and protein content of a plant.

Closely Related Genes from Other Species

G865 and other non-Arabidopsis AP2/EREBP proteins were similar within the conserved AP2 domain.

G2509: G2509 (SEQ ID NO: 23) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Overexpression of G2509 caused multiple alterations in plant growth and development, most notably, altered branching patterns, and a reduction in apical dominance, giving the plants a

shorter, more bushy stature than wild type. Twenty 35S::G2509 primary transformants were examined; at early stages of rosette development, these plants displayed a wild-type phenotype. However, at the switch to flowering, almost all T1 lines showed a marked loss of apical dominance and large numbers of secondary shoots developed from axils of primary rosette leaves. In the most extreme cases, the shoots had very short internodes, giving the inflorescence a very bushy appearance. Such shoots were often very thin and flowers were relatively small and poorly fertile. At later stages, many plants appeared very small and had a low seed yield compared to wild type. In addition to the effects on branching, a substantial number of 35S::G2509 primary transformants also flowered early and had buds visible several days prior to wild type. Similar effects on inflorescence development were noted in each of three T2 populations examined. The branching and plant architecture phenotypes observed in 35S::G2509 lines resemble phenotypes observed for three other AP2/EREBP genes: G865 (SEQ ID NO: 557), G1411 (SEQ ID NO: 3), and G1794 (SEQ ID NO: 13). G2509, G865, and G1411 form a small clade within the large AP2/EREBP family, and G1794, although not belonging to the clade, is one of the AP2/EREBP genes closest to it in the phylogenetic tree. It is thus likely that all these genes share a related function, such as affecting hormone balance. Overexpression of G2509 in Arabidopsis resulted in an increase in alpha-tocopherol in seeds in T2 lines 5 and 11. G2509 was ubiquitously expressed in Arabidopsis plant tissue. G2509 expression levels were altered by a variety of environmental or physiological conditions. G2509 can be used to manipulate plant architecture and development. G2509 can be used to alter tocopherol composition. Tocopherols have anti-oxidant and vitamin E activity. G2509 can be useful in altering flowering time. A wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G2509 showed some sequence similarity with known genes from other plant species within the conserved AP2/EREBP domain.

G2347: G2347 (SEQ ID NO: 1119) was analyzed using transgenic plants in which G2347 was expressed under the control of the 35S promoter. Overexpression of G2347 markedly reduced the time to flowering in Arabidopsis. This phenotype

was apparent in the majority of primary transformants and in all plants from two out of the three T2 lines examined. Under continuous light conditions, 35S::G2347 plants formed flower buds up a week earlier than wild type. Many of the plants were rather small and spindly compared to controls. To demonstrate that overexpression of G2347 could induce flowering under less inductive photoperiods, two T2 lines were re-grown in 12 hour conditions; again, all plants from both lines bolted early, with some initiating flower buds up to two weeks sooner than wild-type. As determined by RT-PCR, G2347 was highly expressed in rosette leaves and flowers, and to much lower levels in embryos and siliques. No expression of G2347 was detected in the other tissues tested. G2347 expression was repressed by cold, and by auxin treatments and by infection by Erysiphe. G2347 is also highly similar to the Arabidopsis protein G2010 (SEQ ID NO: 1121). The level of homology between these two proteins suggested they could have similar, overlapping, or redundant functions in Arabidopsis. In support of this hypothesis, overexpression of both G2010 and G2347 resulted in early flowering phenotypes in transgenic plants.

Closely Related Genes from Other Species

The closest relative to G2347 is the Antirrhinum protein, SBP2 (CAA63061). The similarity between these two proteins is extensive enough to suggest they might have similar functions in a plant.

G988: G988 (SEQ ID NO: 43) was analyzed using transgenic plants in which G988 was expressed under the control of the 35S promoter. Plants overexpressing G988 had multiple morphological phenotypes. The transgenic plants were generally smaller than wild-type plants, had altered leaf, inflorescence and flower development, altered plant architecture, and altered vasculature. In one transgenic line overexpressing G988 (line 23), an increase in the seed glucosinolate M39489 was observed. The phenotype of plants overexpressing G988 was wild-type in all other assays performed. In wild-type plants, G988 was expressed primarily in flower and silique tissue, but was also present at detectable levels in all other tissues tested. Expression of G988 was induced in response to heat treatment, and repressed in response to infection with Erysiphe. Based on the observed morphological phenotypes of the transgenic plants, G988 can be used to create plants with larger flowers. This can have value in the ornamental horticulture industry. The reduction

in the formation of lateral branches suggests that G988 can have utility on the forestry industry. The Arabidopsis plants overexpressing G988 also had reduced fertility. This can be a desirable trait in some instances, as it can be exploited to prevent or minimize the escape of GMO (genetically modified organism) pollen into the environment.

Closely Related Genes from Other Species

The amino acid sequence for the Capsella rubella hypothetical protein represented by GenBank accession number CRU303349 was significantly identical to G988 outside of the SCR conserved domains. The Capsella rubella hypothetical protein is 90% identical to G988 over a stretch of roughly 450 amino acids. Therefore, it is likely that the Capsella rubella gene is an ortholog of G988.

G2346: G2346 (SEQ ID NO: 459) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. 35S::G2346 seedlings from all three T2 populations had slightly larger cotyledons and appeared somewhat more advanced than controls. This indicated that the seedlings developed more rapidly than the control plants. At later stages, however, G2346 overexpressing plants showed no consistent differences from control plants. The phenotype of these transgenic plants was wild-type in all other assays performed. According to RT-PCR analysis, G2346 is expressed ubiquitously.

Closely Related Genes from Other Species

G2346 shows some sequence similarity with known genes from other plant species within the conserved SBP domain.

G1354: The complete sequence of G1354 (SEQ ID NO: 285) was determined. G1354 was analyzed using transgenic plants in which G1354 was expressed under the control of the 35S promoter. Overexpression of G1354 produced highly deleterious effects on growth and development. Only three 35S::G1354 T1 plants were obtained; all were extremely tiny and slow developing. After three weeks of growth, each of the plants comprised a completely disorganized mass of leaves and root that had no clear axis of growth. Since these individuals would not have survived transplantation to soil, they were harvested for RT-PCR analysis; all three plants showed moderate

levels of G1354 overexpression compared to whole wild-type seedlings of an equivalent size. Only a very small number of transformants were obtained from two selection attempts on separate batches of T0 seed. Usually between 15 and 120 transformants are obtained from each aliquot of 300 mg T0 seed from wild-type plants. The low transformation frequency obtained in this experiment suggests that high levels of G1354 overexpression might have completely lethal effects and prevent transformed seeds from germinating. As determined by RT-PCR, G1354 was uniformly expressed in all tissues and under all conditions tested in RT-PCR. However, the gene was repressed in leaf tissue in response to Erysiphe infection.

Closely Related Genes from Other Species

G1354 is closely related to a NAM protein encoded by polynucleotide from rice (AC005310). Similarity between G1354 and this rice protein extends beyond the signature motif of the family to a level that would suggest the genes are orthologs.

G1063: G1063 (SEQ ID NO: 119) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1499 (SEQ ID NO: 7), G2143 (SEQ ID NO: 129), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. A spectrum of developmental alterations was observed amongst 35S::G1063 primary transformants and the majority were markedly small, dark green, and had narrow curled leaves. The most severely affected individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures. In other cases, flowers showed internode elongation between floral whorls, with a central carpel protruding on a pedicel-like organ. Additionally, lateral branches sometimes failed to develop and tiny patches of carpelloid tissue formed at axillary nodes of the inflorescence. In lines with an intermediate phenotype, flowers contained defined whorls of organs, but sepals were converted to carpelloid structures or displayed patches of carpelloid tissue. In contrast, lines with a weak phenotype developed relatively normal flowers and produced a reasonable quantity of seed. Such plants were still distinctly smaller than wild-type controls. Since the strongest 35S::G1063 lines were sterile, three lines

with a relatively weak phenotype, that had produced sufficient seed for biochemical and physiological analysis, were selected for further study. Two of the T2 populations (T2-28,37) were clearly small, darker green and possessed narrow leaves compared to wild type. Plants from one of these populations (T2-28) also produced occasional branches with abnormal flowers like those seen in the T1. The final T2 population (T2-30) displayed a very mild phenotype. Overexpression of G1063 in *Arabidopsis* resulted in a decrease in seed oil content in T2 lines 28 and 37. No altered phenotypes were detected in any of the physiological assays, except that the plants were noted to be somewhat small and produce anthocyanin when grown in Petri plates. G1063 was expressed at low to moderate levels in roots, flowers, rosette leaves, embryos, and germinating seeds, but was not detected in shoots or siliques. It was induced by auxin. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*. G1063 has utility in manipulating seed oil and protein content.

Closely Related Genes from Other Species

G1063 protein shared extensive homology in the basic helix loop helix region with a protein sequence encoded by Glycine max cDNA clone (AW832545) as well as a tomato root, plants pre-anthesis *Lycopersicon esculentum* cDNA (BE451174).

G2143: G2143 (SEQ ID NO: 129) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2557 (SEQ ID NO: 133). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. Twelve out of twenty 35S::G2143 T1 lines showed a very severe phenotype; these plants were markedly small and had narrow, curled, dark-green leaves. Such individuals were completely sterile and formed highly abnormal inflorescences; shoots often terminated in pin-like structures, and flowers were replaced by filamentous carpelloid structures, or a fused mass of carpelloid tissue. Furthermore, lateral branches usually failed to develop, and tiny patches of stigmatic tissue often formed at axillary nodes of the inflorescence. Strongly affected plants displayed the highest levels of transgene expression

(determined by RT-PCR). The remaining T1 lines showed lower levels of G2143 overexpression; these plants were still distinctly smaller than wild type, but had relatively normal inflorescences and produced seed. Since the strongest 35S::G2143 lines were sterile, three lines with a relatively weak phenotype, that had produced sufficient seed for biochemical analysis, were selected for further study. T2-11 plants displayed a very mild phenotype and had somewhat small, narrow, dark green leaves. The other two T2 populations, however, appeared wild-type, suggesting that transgene activity might have been reduced between the generations. Reduced seedling vigor was noted in the physiological assays. G2143 expression was detected at low levels in flowers and siliques, and at higher levels in germinating seed. G2143 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2143 protein shared extensive homology in the basic helix loop helix region with a protein encoded by Glycine max cDNA clones (AW832545, BG726819 and BG154493) and a *Lycopersicon esculentum* cDNA clone (BE451174). There was lower homology outside of the region.

G2557: G2557 (SEQ ID NO: 133) is a member of a clade of highly related HLH/MYC proteins that also includes G779 (SEQ ID NO: 113), G1063 (SEQ ID NO: 119), G1499 (SEQ ID NO: 7), and G2143 (SEQ ID NO: 129). All of these genes caused similar pleiotropic phenotypic effects when overexpressed, the most striking of which was the production of ectopic carpelloid tissue. These genes can be considered key regulators of carpel development. The flowers of 35S::G2557 primary transformants displayed patches of stigmatic papillae on the sepals, and often had rather narrow petals and poorly developed stamens. Additionally, carpels were also occasionally held outside of the flower at the end of an elongated pedicel like structure. As a result of such defects, 35S::G2557 plants often showed very poor fertility and formed small wrinkled siliques. In addition to such floral abnormalities, the majority of primary transformants were also small and darker green in coloration than wild type. Approximately one third of the T1 plants were extremely tiny and completely sterile. Three T1 lines (#7,9,12), that had produced some seeds, and

showed a relatively weak phenotype, were chosen for further study. All three of the T2 populations from these lines contained plants that were distinctly small, had abnormal flowers, and were poorly fertile compared to controls. Stigmatic tissue was not noted on the sepals of plants from these three T2 lines. Another line (#4) that had shown a moderately strong phenotype in the T1 was sown for only morphological analysis in the T2 generation. T2-4 plants were small, dark green, and produced abnormal flowers with ectopic stigmatic tissue on the sepals, as had been seen in the parental plant. G2557 expression was detected at low to moderate levels in all tissues tested except shoots. It was induced by cold, heat, and salt, and repressed by pathogen infection. G1063 can be used to manipulate flower form and structure or plant fertility. One application for manipulation of flower structure can be in the production of saffron, which is derived from the stigmas of *Crocus sativus*.

Closely Related Genes from Other Species

G2557 protein shows extensive sequence similarity in the region of basic helix loop helix with a protein encoded by Glycine max cDNA clone (BE347811).

G2430: The complete sequence of G2430 (SEQ ID NO: 697) was determined. G2430 is a member of the response regulator class of GARP proteins (ARR genes), although one of the two conserved aspartate residues characteristic of response regulators is not present. The second aspartate, the putative phosphorylated site, is retained so G2430 can have response regulator function. G2430 is specifically expressed in embryo and silique tissue. In morphological analyses, plants overexpressing G2430 showed more rapid growth than control plants at early stages, and in two of three lines examined produced large, flat leaves. Early flowering was observed for some lines, but this effect was inconsistent between plantings. G2430 can regulate plant growth. Overexpression of G2430 in *Arabidopsis* also resulted in seedlings that are slightly more tolerant to heat in a germination assay. Seedlings from G2430 overexpressing transgenic plants were slightly greener than the control seedlings under high temperature conditions. In a repeat experiment on individual lines, G2430 line 15 showed the strongest heat tolerant phenotype. G2430 can be useful to promote faster development and reproduction in plants.

Closely Related Genes from Other Species

G2430 had some similarity within of the conserved GARP and response-regulator domains to non-Arabidopsis proteins.

G1478: The sequence of G1478 (SEQ ID NO: 831) was determined and G1478 was analyzed using transgenic plants in which G1478 was expressed under the control of the 35S promoter. Plants overexpressing G1478 had a general delay in progression through the life cycle, in particular a delay in flowering time. G1478 is expressed at higher levels in flowers, rosettes and embryos but otherwise expression is constitutive. Based on the phenotypes produced through G1478 overexpression, G1478 can be used to manipulate the rate at which plants grow, and flowering time.

Closely Related Genes from Other Species

G1478 shows some homology to non-Arabidopsis proteins within the conserved domain.

G681: G681 (SEQ ID NO: 579) was analyzed using transgenic plants in which the gene was expressed under the control of the 35S promoter. Approximately half of the 35S::G681 primary transformants were markedly small and formed narrow leaves compared to controls. These plants often produced thin inflorescence stems, had rather poorly formed flowers with low pollen production, and set few seeds. Three T1 lines with relatively weak phenotypes, which had produced reasonable quantities of seed, were selected for further study. Plants from one of the T2 populations were noted to be slightly small, but otherwise the T2 lines displayed no consistent differences in morphology from controls. In leaves of two of the T2 lines, overexpression of G681 resulted in an increase in the percentage of the glucosinolate M39480. According to RT-PCR analysis, G681 expression was detected at very low levels in flower and rosette leaf tissues. G681 was induced by drought stress. G681 can be used to alter glucosinolate composition in plants. Increases or decreases in specific glucosinolates or total glucosinolate content are desirable depending upon the particular application. For example: (1) Glucosinolates are undesirable components of the oilseeds used in animal feed, since they produce toxic effects. Low-glucosinolate varieties of canola have been developed to combat this problem. (2) Some glucosinolates have anti-cancer activity; thus, increasing the levels or

composition of these compounds might be of interest from a nutraceutical standpoint.

(3) Glucosinolates form part of a plants natural defense against insects. Modification of glucosinolate composition or quantity could therefore afford increased protection from predators. Furthermore, in edible crops, tissue specific promoters can be used to ensure that these compounds accumulate specifically in tissues, such as the epidermis, which are not taken for consumption.

Closely Related Genes from Other Species

G681 shows some sequence similarity with known genes from other plant species within the conserved Myb domain.

G878: G878 (SEQ ID NO: 611) was studied using transgenic plants in which the gene was expressed under the control of the 35S promoter. Analysis of primary transformants revealed that overexpression of G878 delays the onset of flowering in *Arabidopsis*. 11/20 of the 35S::G878 T1 plants flowered approximately one week later than wild type under continuous light conditions. These plants were also darker green, had shorter stems, and senesced later than controls. G878 was ubiquitously expressed. G878 can be used to modify flowering time and senescence, and a wide variety of applications exist for systems that either lengthen or shorten the time to flowering.

Closely Related Genes from Other Species

G878 was highly related to other WRKY proteins from a variety of plant species, such as the *Nicotiana tabacum* DNA-binding protein 2 (WRKY2) (AF096299), and a *Cucumis sativus* SPF1-like DNA-binding protein (L44134).

G374: G374 (SEQ ID NO: 47) was expressed at low levels throughout the plant and was induced by salicylic acid. G374 was investigated using lines carrying a T-DNA insertion in this gene. The T-DNA insertion was approximately three quarters of the way into the protein coding sequence and should result in a null mutation. Homozygosity for a T-DNA insertion within G374 caused lethality at early stages of embryo development. In an initial screen for G374 knockouts, heterozygous plants were identified. Seed from those individuals was sown to soil and eleven plants were PCR-screened to identify homozygotes. No homozygotes were obtained;

6 of the progeny were heterozygous whilst the other 5 were wild type. This raised the prospect that homozygosity for the G374 insertion was lethal. To examine this possibility further, heterozygous KO.G374 plants were re-grown. These individuals looked wild type, but their siliques were examined for seed abnormalities. When green siliques were dissected, around 25% of developing seeds were white or aborted. Embryos from these siliques were cleared using Hoyers solution, and examined under the microscope. It was apparent that embryos from the white seeds had arrested at early (globular or heart) stages of development, whilst embryos from the normal seeds were fully developed. Such arrested or aborted seeds most likely represented homozygotes for the G374 insertion. To support this conclusion, seed was collected from heterozygous plants and sown to kanamycin plates (the T-DNA insertion carried the NPT marker gene). Of the seedlings that germinated, 160 were kanamycin resistant and 107 were kanamycin sensitive. These data more closely fitted a 2:1 (chi-sq., 1df, = 5.5, $0.05 > P > 0.01$) than a 3:1 (chi-sq., 1df, = 32, $P < 0.001$) ratio. Such a segregation ratio suggested that a homozygous class of kanamycin resistant seedlings was absent from the progeny of KO.G374 plant. G374 can be a herbicide target.

Closely Related Genes from Other Species

Similar sequences to G374 are present in tomato and *Medicago truncatula*, and these sequences can be orthologs.

Example VIII: Identification of Homologous Sequences

Homologous sequences from *Arabidopsis* and plant species other than *Arabidopsis* were identified using database sequence search tools, such as the Basic Local Alignment Search Tool (BLAST) (Altschul et al. (1990) *J. Mol. Biol.* 215:403-410; and Altschul et al. (1997) *Nucl. Acid Res.* 25: 3389-3402). The tblastx sequence analysis programs were employed using the BLOSUM-62 scoring matrix (Henikoff, S. and Henikoff, J. G. (1992) *Proc. Natl. Acad. Sci. USA* 89: 10915-10919).

Identified non-*Arabidopsis* sequences homologous to the *Arabidopsis* sequences are provided in Table 5. The percent sequence identity among these sequences can be as low as 47%, or even lower sequence identity. The entire NCBI GenBank database was filtered for sequences from all plants except *Arabidopsis thaliana* by selecting all entries in the NCBI GenBank database associated with NCBI

taxonomic ID 33090 (Viridiplantae; all plants) and excluding entries associated with taxonomic ID 3701 (*Arabidopsis thaliana*). These sequences are compared to sequences representing genes of SEQ IDs NOs:2 - 2N, where $N = 2-561$, using the Washington University TBLASTX algorithm (version 2.0a19MP) at the default settings using gapped alignments with the filter "off". For each gene of SEQ IDs NOs:2 - 2N, where $N = 2-561$, individual comparisons were ordered by probability score (P-value), where the score reflects the probability that a particular alignment occurred by chance. For example, a score of $3.6e-40$ is 3.6×10^{-40} . In addition to P-values, comparisons were also scored by percentage identity. Percentage identity reflects the degree to which two segments of DNA or protein are identical over a particular length. Examples of sequences so identified are presented in Table 5. Homologous or orthologous sequences are readily identified and available in GenBank by Accession number (Table 5; Test sequence ID). The identified homologous polynucleotide and polypeptide sequences and homologues of the *Arabidopsis* polynucleotides and polypeptides may be orthologs of the *Arabidopsis* polynucleotides and polypeptides (TBD: to be determined).

Example IX Introduction of polynucleotides into dicotyledonous plants

SEQ ID NOs:1-(2N - 1), wherein $N = 2-561$, paralogous, orthologous, and homologous sequences recombined into pMEN20 or pMEN65 expression vectors are transformed into a plant for the purpose of modifying plant traits. The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants using most dicot plants (see Weissbach and Weissbach, (1989) *supra*; Gelvin et al., (1990) *supra*; Herrera-Estrella et al. (1983) *supra*; Bevan (1984) *supra*; and Klee (1985) *supra*). Methods for analysis of traits are routine in the art and examples are disclosed above.

Example X Transformation of Cereal Plants with an Expression Vector

Cereal plants such as corn, wheat, rice, sorghum or barley, may also be transformed with the present polynucleotide sequences in pMEN20 or pMEN65 expression vectors for the purpose of modifying plant traits. For example, pMEN020 may be modified to replace the NptII coding region with the BAR gene of *Streptomyces hygroscopicus* that confers resistance to phosphinothricin. The KpnI

and BglII sites of the Bar gene are removed by site-directed mutagenesis with silent codon changes.

The cloning vector may be introduced into a variety of cereal plants by means well-known in the art such as, for example, direct DNA transfer or *Agrobacterium tumefaciens*-mediated transformation. It is now routine to produce transgenic plants of most cereal crops (Vasil, I., Plant Molec. Biol. 25: 925-937 (1994)) such as corn, wheat, rice, sorghum (Cassas, A. et al., Proc. Natl. Acad Sci USA 90: 11212-11216 (1993) and barley (Wan, Y. and Lemeaux, P. Plant Physiol. 104:37-48 (1994). DNA transfer methods such as the microprojectile can be used for corn (Fromm. et al. Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al. Plant Cell 2: 603-618 (1990); Ishida, Y., Nature Biotechnology 14:745-750 (1990)), wheat (Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), rice (Christou Bio/Technology 9:957-962 (1991); Hiei et al. Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617; Hiei et al., Plant Mol Biol. 35:205-18 (1997)). For most cereal plants, embryogenic cells derived from immature scutellum tissues are the preferred cellular targets for transformation (Hiei et al., Plant Mol Biol. 35:205-18 (1997); Vasil, Plant Molec. Biol. 25: 925-937 (1994)).

Vectors according to the present invention may be transformed into corn embryogenic cells derived from immature scutellar tissue by using microprojectile bombardment, with the A188XB73 genotype as the preferred genotype (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). After microprojectile bombardment the tissues are selected on phosphinothricin to identify the transgenic embryogenic cells (Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)). Transgenic plants are regenerated by standard corn regeneration techniques (Fromm, et al., Bio/Technology 8: 833-839 (1990); Gordon-Kamm et al., Plant Cell 2: 603-618 (1990)).

The plasmids prepared as described above can also be used to produce transgenic wheat and rice plants (Christou, Bio/Technology 9:957-962 (1991); Hiei et al., Plant J. 6:271-282 (1994); Aldemita and Hodges, Planta 199:612-617 (1996); Hiei et al., Plant Mol Biol. 35:205-18 (1997)) that coordinately express genes of

interest by following standard transformation protocols known to those skilled in the art for rice and wheat Vasil, et al. Bio/Technology 10:667-674 (1992) ; Vasil et al., Bio/Technology 11:1553-1558 (1993); Weeks et al., Plant Physiol. 102:1077-1084 (1993)), where the bar gene is used as the selectable marker.

All references, publications, patent documents, web pages, and other documents cited or mentioned herein are hereby incorporated by reference in their entirety for all purposes. Although the invention has been described with reference to specific embodiments and examples, it should be understood that one of ordinary skill can make various modifications without departing from the spirit of the invention. The scope of the invention is not limited to the specific embodiments and examples provided.

We claim:

1. A transgenic plant comprising a recombinant polynucleotide having a nucleotide sequence selected from the group consisting of:
 - (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from those of SEQ ID NOs: 860, 802, 240, 274, 558, 24, 1120, 44, 460, 286, 120, 130, 134, 698, 832, 580, 612, and 48, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 859, 801, 239, 273, 557, 23, 1119, 43, 459, 285, 119, 129, 133, 697, 831, 579, 611, 47, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more polynucleotides of: (a) or (b).
2. The transgenic plant of claim 1 wherein the transgenic plant possesses an altered trait as compared to another plant, or the transgenic plant exhibits an altered phenotype as compared to another plant, or the transgenic plant expresses an altered level of one or more genes associated with a plant trait as compared to another plant, wherein the other plant does not comprise the recombinant polynucleotide.
3. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
4. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics,

apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.

5. The transgenic plant of claim 1 wherein the plant possesses an altered trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
6. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:860.
7. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:802.
8. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:240.
9. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:274.
10. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:558.
11. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:24.
12. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:1120.
13. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:44.

14. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:460.
15. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:286.
16. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:120.
17. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:130.
18. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:134.
19. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:698.
20. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
21. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:580.
-
22. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:612.
23. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:48.
24. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:859.

25. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:801.
26. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
27. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
28. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
29. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:23.
30. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:1119.
31. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:43.
32. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:459.
33. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
34. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:119.
35. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:129.

36. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:133.
37. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
38. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
39. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:579.
40. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:611.
41. The transgenic plant of claim 1, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:47.
42. The transgenic plant of claim 1, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to said nucleotide sequence.
43. The transgenic plant of claim 1, wherein the plant is selected from the group consisting of: soybean, wheat, corn, potato, cotton, rice, oilseed rape, sunflower, alfalfa, sugarcane, turf, banana, blackberry, blueberry, strawberry, raspberry, cantaloupe, carrot, cauliflower, coffee, cucumber, eggplant, grapes, honeydew, lettuce, mango, melon, onion, papaya, peas, peppers, pineapple, pumpkin, spinach, squash, sweet corn, tobacco, tomato, watermelon, mint and other labiates, rosaceous fruits, and vegetable brassicas.
44. The transgenic plant of claim 1 wherein the encoded polypeptide is expressed and regulates transcription of a gene.

45. A method of using the transgenic plant of claim 1 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.
46. An isolated or recombinant polynucleotide comprising a nucleotide sequence selected from the group consisting of:
- (a) a nucleotide sequence encoding a polypeptide comprising a sequence selected from SEQ ID NOs: 240, 274, 558, 286, 698, and 832, or a complementary nucleotide sequence thereof;
 - (b) a nucleotide sequence of SEQ ID NOs: 239, 273, 557, 285, 697, 831, or a complementary nucleotide sequence thereof; and
 - (c) a nucleotide sequence that hybridizes under stringent conditions to a nucleotide sequence of one or more of: (a) or (b).
47. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 240.
48. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 274.
49. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 558.
50. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 286.
51. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO: 698.

52. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence encoding a polypeptide of SEQ ID NO:832.
53. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:239.
54. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:273.
55. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:557.
56. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:285.
57. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:697.
58. The isolated or recombinant polynucleotide of claim 46, wherein the recombinant polynucleotide comprises a nucleotide sequence of SEQ ID NO:831.
59. The isolated or recombinant polynucleotide of claim 46, further comprising a constitutive, inducible, or tissue-specific promoter operably linked to the nucleotide sequence.
60. The isolated or recombinant polynucleotide of claim 46 wherein the encoded polypeptide is expressed and regulates transcription of a gene.
61. A vector comprising the isolated or recombinant polynucleotide of claim 46.
62. A host cell comprising the vector of claim 61.

63. A method of using the isolated or recombinant polynucleotide of claim 46 for producing a plant having a modified trait, the method comprising selecting a polynucleotide that encodes a polypeptide, inserting the polynucleotide into an expression vector, introducing the vector into a plant or a cell of a plant to overexpress the polypeptide, thereby producing a modified plant, and selecting a modified plant for a modified trait.
64. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more physical characteristics selected from the group consisting of: the number of trichomes, fruit and seed size and number, yield of stems, leaves, inflorescences, or roots, stability of seeds during storage, susceptibility of the seed to shattering, root hair length and quantity, internode distances, or the quality of seed coat.
65. The method of claim 63 wherein the plant possesses a modified as compared to another plant wherein the trait is an alteration in a plant growth characteristic selected from the group consisting of: growth rate, germination rate of seeds, vigor of plants and seedlings, leaf and flower senescence, male sterility, apomixis, flowering time, flower abscission, rate of nitrogen uptake, osmotic sensitivity to soluble sugar concentrations, biomass or transpiration characteristics, apical dominance, branching patterns, number of organs, organ identity, and organ shape or size.
66. The method of claim 63 wherein the plant possesses a modified trait as compared to another plant wherein the trait is an alteration in one or more characteristics selected from the group consisting of protein or oil production, seed protein or oil production, insoluble sugar level, soluble sugar level, and starch composition.
67. A modified plant produced by the method of claim 63.
68. A method of using the plant of claim 67 to grow a progeny plant from a parent plant, the method comprising crossing the transgenic plant with another plant, selecting seed, and growing the progeny plant from the seed.

69. The plant produced by the method of claim 68.

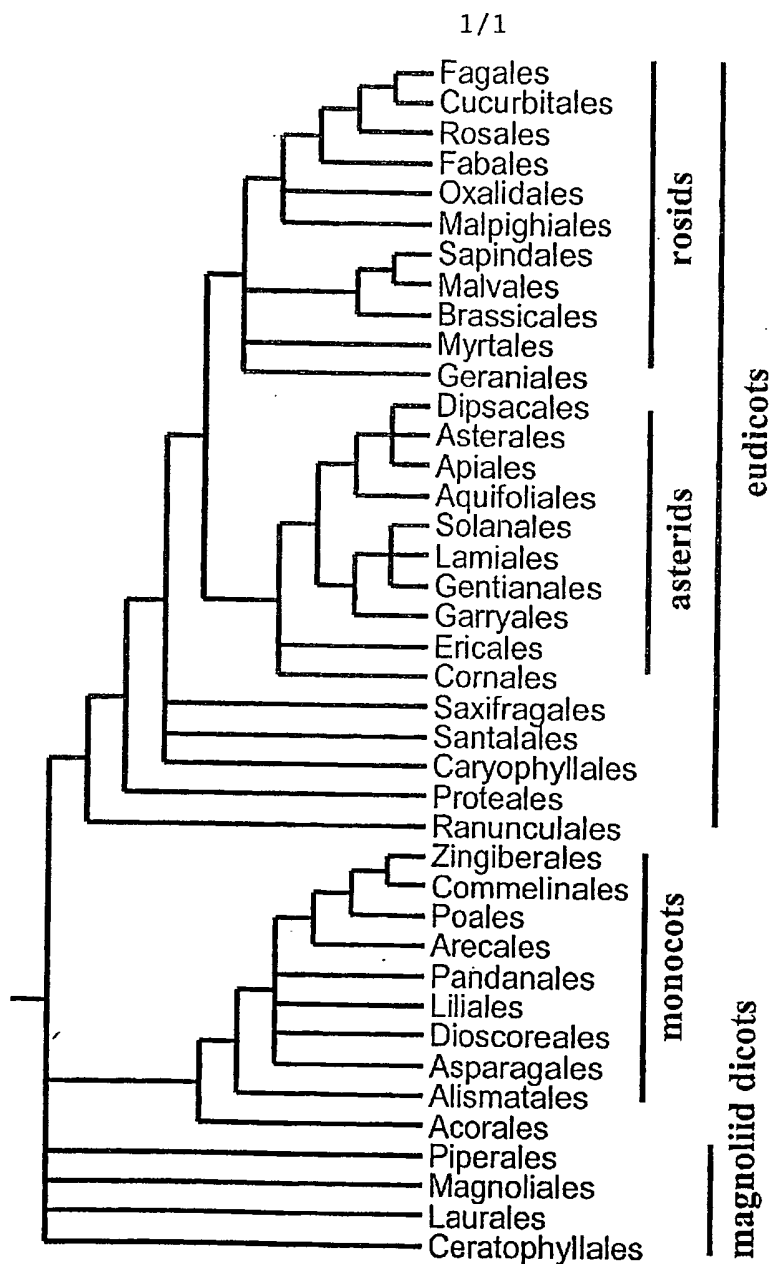


Figure 1

SEQUENCE LISTING

<110> Mendel Biotechnology, Inc.

Ratcliffe, Oliver
 Riechmann, Jose Luis
 Adam, Luc J.
 Dubell, Arnold T.
 Heard, Jacqueline E.
 Pilgrim, Marsha L.
 Jiang, Cai-Zhong
 Reuber, T. Lynne
 Creelman, Robert A.
 Pineda, Omaira
 Yu, Guo-Liang
 Broun, Pierre E.

<120> YIELD-RELATED POLYNUCLEOTIDES AND
POLYPEPTIDES IN PLANTS

<130> 514442002041

<150> 60/310,847

<151> 2001-08-09

<150> 60/336,049

<151> 2001-11-19

<150> 60/338,692

<151> 2001-12-11

<150> 10/171,468

<151> 2002-06-14

>G1275 (58..579)

CCAAGAAAAGGGAAGATCACGCATTCTTATAGGCGTAATTCGTAAATAGTGGTGAGTATG
 AATGATGCAGACACAACTTGGGGAGTAGTTTCAGCGATGATACTCACTCTGTGTTTCGAG
 TTTCCGAGCTAGACTTGTCTAGATGAATGGATGGATGATGATCTTGTGTCTGCGGTTTCC
 GGGATGAATCAGTCTTATGGTTATCAGACTAGTGATGTTGCTGGTGCTTTATTCTCAGGT
 TCTTCTAGCTGTTTCACTCATCTGAATCTCCAAGTACCAAACTTATGTTGCTGCTACA
 GCCACTGTTCTGCCGACAACCAAAACAAGAAAAGAAAAGAAAAAATTAAAGGGAGAGTT
 GCGTTCAAGACACGGTCCGAGGTGGAAGTGCTTGACGACGGGTTCAAGTGGAGAAAGTAT
 GGAAGAAGATGGTGAAGAACAGCCACATCCAAGAACTACTACAAATGTTTCAAGTTGAT
 GGCTGTCCCGTGAAGAAAAGGGTTGAACGAGACAGAGATGATCCGAGCTTTGTGATAACA
 ACTTACGAGGGTTCCCAATCACTCAAGCATGAACCTAAGACTCGAACTAAGGCTCAAGG
 CGACCATGCTATATTCAGCACATCTTATTTCTATGGTTACGAACGATACTTAAACTGC
 TTCTAGTTCTTTATATCCATTGTAACTGGTTGCAGGTTACAAAATTTTGAGAGGTTTAT
 GACATTCTAAATCTGTAGTACTTATATA

>G1275 Amino Acid Sequence (domain in AA coordinates: 113-169)

MNDADTNLGSSFSDDTHSVFEFPELDLSDEWMDDDLVSAVSGMNQSYGYQTSVAGALFS
 GSSSCFSPHSPSTKTYVAATATASADNQNKKEKKIKGRVAFKTRSEVEVLDDGFKWRK
 YGKMKVKNSPHPRNYKCSVDGCPVKRVERDRDDPSFVITTYEGSHNHSSMN*

>G1411 (110..856)

TAAAGAAAACTGAACAACCCTAAAGTACTGTATAAATCCTATATCAAATTTTTTTTTTG
 GAAGAAAAGGCTATATTTAAAGAAAATCAAGCAAAAGTAGATCCTCGGATGTATGGGAA
 GAGGCCTTTGGAGGTGATGAATCTGAAGAAAGGGAAGAAGATGAGAAGTTGTTCCCGGT
 CTTCTCGGCCCGATCTCAACACGACATGCGTGTTATGGTCTCGGCCCTGACTCAAGTAAT

CGGAAACCAACAAAGCAAATCTCATGATAACATCAGCTCTATTGATGATAACTATCCTTC
 TGTGTATAATCCACAAGACCCTAATCAACAAGTTGCGCCTACTCATCAAGACCAAGGGGA
 CTTGAGGAGGAGACATTATAGAGGTGTAAGGCAAAGGCCATGGGGAAGTGGGCAGCTGA
 AATCCGAGACCCAAAAAAGGCGGCACGTGTGTGGCTCGGGACATTGAAACCGCTGAATC
 TGGCGCCTTAGCTTATGATGAAGCAGCCCTAAAGTTCAAAGGAAGCAAAGCAAACTCAA
 TTTCCCGGAGAGGGTTTCAGCTTGGAGTAACCTACATATTACTCCTCCAACCAAAATCC
 ACAAATGGAACCACAAAGTATACCGAACTATAATCAATACTATCATGATGCGAGTAGTGG
 TGATATGCTAAGTTTTAATTTGGGCGGTGGGTATGGGAGTGGTACCGGATATTCAATGTC
 TCATGATAATAGTACTACGACTGCTGCTACAACCTTCTTCGTCTTCTGGTGGCTCTTCTAG
 GCAACAAGAAGAGCAAGATTATGCCAGATTCTGGCGCTTTGGGGATTCTTCTTCTCTCC
 TCATTCGGGATATTAAATTAGGAGATTGATCAGTTACTTGTGATGAAGTAATGATACATT
 TCCCGTCAAAATTGAGATGATCATATGCTTCTGAAATGTTTTTGTAGTGTCATTTTTGTCT
 TCCCGCTTAAGATTTATGAACGTGTTTTCTTGTTTTTTGGTTAAAAA

AAAAA
 >G1411 Amino Acid Sequence (domain in AA coordinates: 87-154)
 MYGKRPFPGDESEEREEDENLFPVFSARSQHDMRVMVSALTQVIGNQQSKSHDNISSIDD
 NYPSVYNPDQPNQOVAPTHQDQGLRRRHRYRGVRQRPWGKWAAEIRDPKKAARVWLGTFE
 TAESAALAYDEAALKFKGSKAKLNFPERVQLGSNSTYSSNQIPQMEPQSIPNYNQYHYD
 ASSGDMLSFNLGGYGS GTGYSMSHDNSTTAATTSSSSGGSSRQEEQDYARFWRFGDS
 SSSPHSGY*

>G1488 (1..996)
 ATGGAAGATGAAGCACATGAATCTTCCACACATCTGATTTTGCCGTTGATGACCTTTTA
 GTTGATTTCTTAACGATGATGACGAAGAAACGATGTTGTTGCTGATTCACCACTACC
 ACCACCATAACCGACAGCTCTAACTTCTCCGCTGCTGATCTTCCAGTTTCCACGGTGAT
 GTTCAAGACGGCACTAGCTTCTCCGGTGACCTTTGTATACCTTCTGATGATTTGGCTGAT
 GAGTTAGAGTGGCTTTTCGAACATTTGTGGATGAATCATTGTGCGCTGAAGATGTACACAAG
 CTCGAGCTAATATCCGGTTTTAAGAGTTCGACCGGACCCGAAATCCGATACCGGAAGCCCG
 GAAAACCCGAATAGCAGCAGTCCGATTTTTACTACCGACGTTTCTGTACCGGCAAAGCT
 AGAAGCAAACGCTCACGCGCCGCTGCGTGTAATTGGGCCTCACGTGGGCTTCTCAAGGAA
 ACGTTTTACGACAGTCCCTTTCACCGGAGAAACATTCTCTCTAGCCAACAACACTTGTCT
 CCGCCAACCTCGCCGCTTTGTTGATGGCTCCGCTAGGGAAAAAGCAAGCCGTTGATGGA
 GGACACCGACGGAAGAAGGATGTTCTTACCGGAGTCTGGTGGCGCAGAGGAGAGACGG
 TGTCTCCACTGCGCCACGGATAAGACTCCGCAATGGCGGACAGGCCCAATGGGCCGAAG
 ACGTTGTGCAACGCTTGGGTTGTAGGTACAAATCGGGACGTTTAGTGCCGGAGTATCGG
 CCCGCGGCGAGTCCGACGTTTGTGCTGGCGAAACACTCAAATCTCATCGGAAAGTTATG
 GAGCTCCGCGACAGAAGGAGATGAGTAGGGCCCATCATGAGTTCATACATCACCATCAC
 GGTACGGACACTGCCATGATTTTCGACGTTTCATCGGACGGTGATGATTACTTGATCCAC
 CACAACGTTGGCCAGATTTTCAGACAGCTTATTTGA

>G1488 Amino Acid Sequence (domain in AA coordinates: 221-246)
 MEDEAHEFFHTSDFAVDDLLVDFSNDDDEENDVVADSTTTTTITDSSNFSAADLPSFHGD
 VQDGTFSFGDLCLIPSDDLADELEWLSNIVDESLSPELVHKLLEISGFKSRPDPKSDTGSP
 ENPNSSSPIFTTDSVPAKARSKRSRAAACNWSRGLLKETFYDSPFTGETILSSQQHLS
 PPTSPPLLMAPLGKKQAVDGGHRRKKDVSSPESGGAERRLHCATDKTPQWRTGPMGPK
 TLCNACGVRYKSGRLVPEYRPAASPTFVLAKHSNSHRKVMELRRQKEMSRAHHEFIHHHH
 GTDTAMIFDVSSDGDYLIHNVGPDFRQLI*

>G1499 (159..833)
 TCGACTCCTTAATTGCATCACCAACCTAACCCCTCATCATATGATCACCTTTCTTTTTT
 CCTTTAATATATATATATGCTCACACACACATATATATATACATATAAGCATCGCCTC
 AAGCATTAAAATTTTACGAACCAAAACAAAATATGAATAATTATAATATGAACC
 CATCTCTCTTCCAAAATTACACTTGGACAACATCATCAACAGCAGCAACAACAACA
 AGAATGATGATCATCATCAACATAATAATGATCCAATCGGTATGGCCATGGACCACT
 ACACACAGCTCCATATCTTCAATCTTTCTCTTCTTCTCATTTCCTCTCTCTCTCTT
 CCCTCACAAACCACACTCTTCTCTCCGAGATCAAGAAGACGACGAAGACGAAGAAGAAC
 CTCTAGAGGAACCTCGGTGCTATGAAGGAATGATGTACAAGATCGCAGCCATGCAATCGG
 TTGACATCGACCCAGCAACCGTCAAGAAACCCAAACGCCGTAACGTGAGGATCTCCGACG
 ACCCTCAGAGTGTGGCGGCTAGACATCGCCGTGAGAGAATCAGTGAGAGGATCAGAAATC
 TTCAGAGACTCGTGCCAGGTGGCACTAAATGGATACGGCTTCAATGCTCGATGAAGCTA

TACGCTATGTCAAGTTCTTGAAACGGCAGATCCGGCTACTCAATAATAATACCGGATATA
CTCCTCCGCCGCCGCAAGATCAAGCTTCTCAGGCGGTGACGACGTCATGGGTTTCACCGC
CACCACCGCCAAGTTTCGGCCGTGGGGGAAGAGGAGTAGGAGAATTAATCTAGACAAGAT
GACATTTCCATTAGTAGTAATAAATTATGCTATAATGTGTGAGTAATGGTGCAATTATG
GA

>G1499 Amino Acid Sequence (domain in AA coordinates: 118-181)
MNNYNMNP SLFQNYTWNNIINSSNNNNKNDH HHQHNNDP IGMAMDQYTLHIFNPFSSS
HFPPLSSSLT TLLSGDQEDDEDEEEPLEELGAMKEMMYKIAAMQSVDIDPATVKKPKR
RNVRI SDDPQSVAAHRHRRERISERIRILQRLVPGG TKMDTASMLDEAIRYVKFLKRQIRL
LNNNTGYTPPP PQDQASQAVTTSWVSPPPPPSFGRGGRGVGELI*

>G1543 (1..828)

ATGATAAACTACTATTTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAATAAAGGCATCGTCACGCTTCAAGTT
TGTCCTCTTTTATATTA AACTGAGAGTTTTCCTCTCAAAC TTTACCTTTTCTTCTCGATC
CTAGCTCTTAAGAACCCTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGGA AAAAC
TCTTCAAAC TTTGGATCTTACTATCTCCGTTCCAGGCTTCTCTTCATCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAGTCACGATGATGGCTCTGCTCCTCCGCGAAAGAAA
CTCCGCTCTAACCAGAGAACAGTCACGCTCTTCTGAAGATAGTTTTCAGACAGAATCATACC
CTTAATCCCAAACAAAAGGAAGTACTTGCCAAGCATTGATGCTACGGCCAAGACAAATT
GAAGTTTGTTTCAAACCGTAGAGCAAGGAGCAAATTGAAGCAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTTCAATTAACGGAAGAAAACCAAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTGCGCTCGAGCCTT
ACTATGTGTCTCTCGCTGCGAGCGAGTTACCCCTGCCGCGAGCCCTTCGAGGGCGGTGGTG
CCGGTCCCGCTAAGAAAACGTTTCCGCCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)
MIKLFLTYICTYTYKLYALYHMDYACVCMYKYKGI VTLQVCLFYIKLRVFLSNFTFSSSI
LALKNPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDESGSGGRDQLRLDMNRLPSS
EDGDDEEF SHDDGSAPPRKKRLRLTREQSRLL EDSFRQNHTLNPQKEVLAKHLMRLPRQI
EVWFQNRARRSKLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKKTFFPPQERDR*

>G1635 (1..1164)

ATGGCGTCTGCTCCGTTGACTGCAAATGTTT CAGGGTACCAACGCTTCTTTGAGGAATAGA
GATGAAGAACTGCAGACAAGCAGATACAATTCAATGACCAAAGTTTGGGGGAAATGAC
TATGCACCCAAGGTACGGAAGCCATACACGATAACAAAAGAGAGAGAGATGGACAGAT
GAAGAGCACAAGAAGTTTGTGTAAGCCTTGAAATTATACGGGCGAGCTTGGAGACGAATA
GAAGAACATGTGGGCTCAAAGACCGCAGTTCAGATTGCAAGCCATGCTCAGAAAGTTTTTC
TCTAAGGTTGCTCGAGAACCACTGGAGGTGATGGGAGCTCAGTAGAGCCGATTGTAATA
CCTCCTCCTCGTCCCAAGAGAAAAGCCAGCGCATCCGTACCTCGTAAGTTTGGGAACGAG
GCAGATCAAACAAGTAGATCGGTTTCTCCCTCAGAACGTGATACTCAATCTCCAACCTCT
GTGTTGTCTCACTGTTGGATCAGAAGCATTGTGTTCCCTTGATTGAGTTACCCCAATCGA
AGCTTGTCCTCCAGTTTCTTCTGCATCACCACCAGCTGCTCTTACAACCACTGCAAATGCA
CCTGAAGAGCTTGAGACTCTGAAGCTGGAGTTGTTTCTAGTGAGAGACTCTTAAACAGG
GAGAGCTCGATCAAGGAACCAACGAAGCAAAGTCTTAAACTCTTTGGGAAGACAGTTTTG
GTATCTGATT CAGGCATGTCTCTTCTCTAACAACCTTCAACATATTGTAAATCCCAATT
CAGCCATTACCAGGAAACTCTCATCATCCAAGACACTACCCATAATAAGAAACTCACAA
GAAGAAGCTCTGAGCTGTGGATACAAAGTCCCTCTTAAGCAAGAAGATGTGGAAAATAGA
TGTTTGGATT CAGGAAAGGCTGTCCAAAACGAAGGATCATCGACTGGATCAAACACTGGT
TCGGTGGATGATACGGGACACACGGAAGACCAAGAACCCGAAACAATGCTATGTCAA
TGGGAGTTTAAACCAAGTGAGAGGCTGTCATTTTCTGAGCTCAGAAGAACAACTCCGAG
TCAAATTCAGAGGATTGGTCCATACAAGAAGAGAAAGATGGTAACAGAAGAAGAGAG
CATGAGATTCATCTCCACTTATAA

>G1635 Amino Acid Sequence (domain in AA coordinates: 44-104)
MASSPLTANVQGTNASLRNDEETADKQIQFNDQSFSGNDYAPKVRKPYTITKERERWTD
EEHKKFVEALKLYGRAWRRIEEHVGSKTAVQIRSHAQKFFSKVAREATGGDGSSVEPIVI
PPRPKRKPAHPYPRKFGNEADQTSRSVSPSERDTQSPTS SVLSTVGSEALCSLDSSSPNR
SLSPVSSASPPAALTTTANAPEELET LKLELFP SERLLNRESSI KEPTKQSLKLFKTVL

VSDSGMSSSLTTSTYCKSPIQPLPRKLSSSKTLPIIRNSQEELLSCWIVPLKQEDVENR
CLDSGKAVQNEGSSTGSNTGSVDDTGHTTEKTTEPETMLCQWEFKPSERSAFSELRRTNSE
SNSRFGFPYKKRKMVTEEEHEIHLHL*

>G1794 (160..1335)

TCTTTCTTTCTTCTCTTTGTCTCTGTTTCTGTTTCTCTCTCTCTCTCTACAGAGTT
TTCTTTCCCTCGAAGAAAAAGAATATTTTAAATTTAATTTCTCTGCGTTTATAAGCTT
TAAGTTTCAGAGGAGGAGGATTTAGAAGGAGGTTTTGTATGTGTCTTAAAAGTGGCA
AATCAGGAAGATAACGTTGGCAAAAAGCCGAGTCTATTAGAGACGATGATCATCGGACG
TTATCTGAAATCGATCAATGGCTTTACTTATTCGCAGCCGAAGACGACCACCCTCAT
AGCTTCCCTACGCAGCAGCCGCTCCATCGTCGTCGTCCTCATCTCTTATCTCAGGTTTC
AGTAGAGAGATGGAGATGTCTGCTATGTCTCTGCTTTGACTCACGTTGTTGCTGGAAAT
GTTCTCTCAGCATCAACAAGGAGGCGTGAAGGTAGCGGAGAAGGACTTCGAATTCGTCT
TCTTCTCTCGGGGAGAAAAGGAGGAGAGAGGTGGAGGAAGGTGGCGCCAAAGCGTTAAG
GCAGCTAATACTTTGACGGTTGATCAATATTTCTCCGGTGGTAGCTCTACTTCTAAAGTG
AGAGAAGCTTCGAGTAACATGTCTGAGTCCGGGCCCAACATACGAGTATACAACCTACGGCA
ACTGCTAGTAGCGAAACGTCGTCTGTTAGTGGGGACCAACCTCGGCGAAGATACAGAGGA
GTTAGACAAAGACCATGGGGAAAGTGGGCGGCTGAGATTCGAGATCCATTTAAAGCAGCT
AGAGTTTGGCTCGGTACGTTTCGACAATGCTGAATCAGCAGCAAGAGCTTACGACGAAGCT
GCACTTCGGTTTAGAGGCAACAAGCCAACTCAACTTCCCTGAAAACGTCAAACTCGTT
AGACCTGCTTCAACCGAAGCACAACCTGTGCACCAACCGCTGCTCAAAGACCGACCCAG
TCAAGGAACCTCGGTTCAACGACTACCTTTTGCCCATAGACCTGCTTCGAATCAAAAGC
GTTTATTTCGCAGCCGTTGATGCAATCATACAACCTTGAGTTACTCTGAAATGGCTCGTCAA
CAACAACAGTTTCAGCAACATCATCAACAATCTTTGGATTATACGATCAATGTCTGTTT
CCGTTGCGTTTTCGGTCACTGGAGGTTCAATGATGCAATCTACGTCGTCTCATCATCTCT
CATTCTCGTCTCTGTTTTCCCGGCTGCTGTTTTCAGCCGCCACCAGAATCAGCTAGCGAA
ACCGGTTATCTCCAGGATATACAAATGGCCATCAGACAAGACTAGTAATACTACAATAAT
AGTCCATCCTCCTGATGACTTGCTTCATTTTATTTGTTTCTACTATAGAGTAATAGAAAAC
AGGAAAATGATTATATGTTATAGAGTTATTTTCCAAATATTATAGGGTTTAGGTTGTTT
GTATTGTTCTGCTTTCATCCTCTCATGCTTTTTTTCTTAATTTATTATTTTGCATTA
TAATTTCTGTTTCATTGTAACAAACATTAAAAAGACCACATGGAGAAAGGAAAAAAGAG
AG

>G1794 Amino Acid Sequence (domain in AA coordinates: TBD)

MCVLKVANQEDNVGKKAESIRDDHRTLSEIDQWLYLFAAEDDHRHSFPTQQPPSSSS
SSLISGFSREMESAIVSALTHVAGNVPHQHQGGEGSGEGTSNSSSSSGQKRREVEE
GGAKAVKAANTLTVDQYFSGGSSTSKVREASSNMSGPGPTYEYTTTATASSETSSFSGDQ
PRRRYRGVRQRPWGKWAIEIRDPFKAAARVWLGTFDNAESAARAYDEAALRFRGNKAKLNF
PENVKLVRPASTEAPVHQTAAQRPTQSRNSGSTTTLLPIRPASNQSVHSQPLMQSYNLS
YSEMARQQQFQHHQQLDLYDQMSFPLRFHGTGGSMQSTSSSSSHSRPLFSPAAPVQ
PPESASETGYLDIQWPSDKTSNNYNNSPSS*

>G1839 (38..592)

ATCACAGTTATGTTTCCATTTCATTGGCTATAAAAACCATGCTCACTCCCTTTTGTCTTC
ACACCATTTCAGGAGAAAAATGAATAGTTGTCTAGTCTAATCCCACCAAAATGGATAATTC
AGAAAATGTTCTATTTAATGATCAAAACGAAAATTTCACTTGTTCACCAACACCTTC
TTCTTCGTACTTGACAAGAGATCAAGAGCACGAGATCATGGTCTCTGCTCTGCGACAAGT
GATATCTAACTCCGGAGCTGACGACGCTCATCAAACTTGATCATCACAAGCGTTCC
GCCTCCAGACGCTGGCCCTTGCTCTCTGTTGGCGTCGCCGTTGCTACGGCTGCACATT
ACAACGGCCGACCGAGAGGTAAAGAAGGAGAAGAAATACAAGGAGTAAGGAAAAAACC
ATCGGGTAAATGGGCGGCGGAGATATGGGATCCGAGATCGAAATCAAGGAGGTGGCTTGG
AACGTTTCTTACGGCGGAGATGGCGGCACAATCTTACAATGATGCGGCGGCTGAGTATCG
AGCAAGACGTGGTAAACAAACGGAGAAGGAATTAACGGCGGTGGAGATGACTGAGAAG
GACATGGTGGTGATCATACACGGCGAGGTGGAAATGTTATATTACTATTGAAAATAA
ATTATTTATTATAGAGGGAGATATTACTCTTTACGCTTTTCATTAAAGATTATTTTATAA
GTTTTAAAGTATTTTATTGTTATAAAAAA

>G1839 Amino Acid Sequence (domain in AA coordinates: TBD)

MLTPFCSSHLQEKMNCSQSNPTKMDNSENVLFDQNFNFTLVAPHPSSSYLTRDQEHEI
MVSALRQVISNSGADDASSNLIITSVPPPDAGPCPLCGVAGCYGCTLQRPHEVKEKK
YKGVRRKPSGKWAIEIWDPRKSRRWLGTFLTAEMAAQSYNDAAEYRARRGKTNGEGIK

RRWR*

>G2108 (35..694)

GAGAGAGAAACATTGATCTCTGAATATTGTGAACATGTTGAAATCAAGTAACAAGAGAAA
AAGCAAAGAAGAGAAGAAGTTACAAGAAGGGAAGTACCTTGGAGTGAGGAGACGTCCATG
GGGAAGATATGCAGCTGAAATCAGAAACCCTTTTACTAAAGAAAGACATTGGCTTGAAC
GTTTGATACAGCCGAAGAAGCTGCTTTTGCATATGACGTTGCTGCTCGATCCATCAGCGG
CTCTCTAGCTACAAACAACTTCTTCTACACTGAAACACCTCTTTAGAAAGACATCCACA
ACAGTCTTTGGAGCCTCATATGACTTGGGGATCTTCTAGTCTCTGCTTCTTCAAGATCA
GCCTTTTGAAAACAACCATTTTGTGCTGATCCTATCTCTTCTTCTTTTCTCAAAAACA
AGAGTCTTCTACCAATCTCACTAACACTTTCTCACATGTTTATAATGATGGTGATCATGT
TGGCCAAAGCAAAGAGATTTCTTTACCTAATGATATGTCAAACAGTTTATTTCGGTCATCA
GGACAAAGTCGGTGAACATGACAATGCAGACCATATGAAGTTTGGCTCAGTTCTCAGCGA
CGAACCTCTCTGCTTTGAGTATGACTACATTGGGAATTATCTTCAGAGTTTCTCAAAGA
TGTCACGACGATGCTCCACAGTTTCTTATGTGAGCTTGTATTACCGATCCTTCAATTTA
TG

>G2108 Amino Acid Sequence (domain in AA coordinates: 18-85)
MLKSSNKRKSKEEKLQEGKYLGVRRRPWGRYAAEIRNPFTKERHWLGTFTAEAAAFAY
DVAARSISGSLATTNFFYTENTSLEHHPQQSLEPHMTWGSSSLCLLQDQPFENNHFVADP
ISSFSQKQESSTNLNTNFSHCYNDGDHVGQSKEISLPNDMSNSLFGHQDKVGEHDNADH
MKFGSVLSDEPLCFEYDYIGNYLQSFLKDVNDAPQFLM*

>G2291 (27..797)

GCTTTCTCACCTTTATAAAATAGAAAATGGAAAACAGCTACACCGTTGATGGTCACCGTC
TTCAATATTCCGTTCCGTTAAGCTCCATGCATGAAACCAGTCAAAACTCCGAAACTTACG
GATTATCCAAAGAGTCGCCGTTGGTCTGCATGCCCTTGTTCGAAACCAACTACTTCAT
TCGATATCTCTCTCTTTTCTCGTTTAAACCAAAACCAGAACCCGAAACACGCATCGTG
TCATGGACGATTCCATCGCCGCCGTCGTGGGCGAAAACGTTCTTTTCGGTGATAAAAACA
AAGTCTCTGATCACTTGACCAAAGAAGGTGGTGTGAAGCGGGGGCGAAGATGCCGCAGA
AGACCGGAGGATTCATGGGAGTGAGAAAACGGCCGTGGGGGAGATGGTCGGCGGAGATAA
GAGACAGGATAGGGCGGTGCAGACACTGGTTAGGAACGTTTCGACACGGCGAAGAGGCAG
CGCGTGCGTATGACGCGGCGGCGAGGAGGCTTAGAGGGACCAAAGCCAAGACCAATTTTCG
TGATTCCTCCGCTTTTCCCAAGGAAATAGCTCAGGCTCAGGAGGATAATAGGATGAGGC
AGAAGCAGAAGAAGAAGAAGAAAAAGTGAGTGTGAGGAAGTGTGTTAAAGTCACAT
CGGTTGCACAGTTGTTCGATGATGCCAATTTTATAAATTCTTCTAGTATTAAAGGAAATG
TGATTAGTTCTATTGATAATCTTGAAAAAATGGGTCTAGAGCTTGATTTGAGTTAGGGT
TGTTGTCTAGGAAGTGATAAAGCACTCGTAGTTAAGTAGTTGTAGTT

>G2291 Amino Acid Sequence (domain in AA coordinates: TBD)
MENSYTVDGHRLQYSVPLSSMHETSQNSETYGLSKESPLVCMPLFETNTTSFDDISSLFSS
NPKPEPENTHRVMDSDIAAVVGENVLFGDKNKVSDHLTKEGGVKRGRKMPQKTGGFMGVR
KRPWGRWSAEIRDRIGRCHWLGTFTAEBAARAYDAAARRLRGKAKTNFVIPPFPKE
IAQAQEDNMRMQKQKKKKKKVSRKCVKVTSAQLFDDANFINSSSIKGNVISSIDNLE
KMGLELDLSLGLLSRK*

>G2452 (1..804)

ATGTCATCGTCGACGATGTACAGAGGAGTTAATATGTTTTCACCGCAAACACAACTGG
ATTTTTCAAGAAGTCAGAGAAGCCACGTGGACGGCGGAGGAGAACAACCGTTTCGAGAAA
GCTCTCGCTTATCTGGACGACAAAGACAATCTTGAGAGCTGGTCCAAGATCGCAGATTG
ATTCGCCGCAAAACAGTAGCTGACGTCATTAAACGATACAAGGAGCTAGAGGATGATGTC
AGCGACATCGAAGCCGGACTTATCCCATTCGGGATACGGCGGCGACGCCTCTCCGCT
GCAAACAGTGACTATTTCTTTGGTCTAGAAAACTCCAGCTACGGTTATGATTACGTCGTT
GGAGGAAAGAGGAGTTCCGCCGGCATGACTGATTGTTTTAGGTCTCCGATGCCGGAAG
GAGAGGAAGAAAGGAGTTCCGTGGACCGAGGACGAACACCTACGATTTCTGATGGGTTTG
AAGAAATATGAAAAGGAGATTGGAGAAACATAGCAAAAAGCTTTGTGACGACTCGAACG
CCGACGCAAGTCGCTTACACGCTCAGAAATATTTTCTTCGACAACCTACAGATGGTAAA
GACAAAAGACGATCAAGTATTCAGGATATCACCAGTGTAAACATCCCTGACGCAGACGCA
TCCGCAACCGCCACGACCGCTGACGTAGCACTCTCTCCTACTCCAGCCAATTTCTTTGAC
GTTTCTCTTCAGCCAATCTCATTACAGTTTCGCGTCTGCGTCTAGCTATTAT
AATGCGTTTTCCGACAGTGGAGTTAA

>G2452 Amino Acid Sequence (conserved domain in AA coordinates: 27-213)

MSSSTMYRGVNMFS PANTNWI FQEVREATWTA EENKRF EKALAYLDDKDNLESWSKIADL
IPGKTVADVIKRYKELEDDVSDIEAGLIPIPGYGGDASSAANS DYFFGLENS SYGYDYV
GGKRSSPAMTDCFRSPMPEKERKKGVPTWTEDEHLRFLMGLKKYKGDWRNIAKSFVTTTRT
PTQVASHAQKYFLRQLTDGKDKRRSSIHDITTVNI PDADASATATTADVALSPTPANSFD
VFLQPNPHYSFASASASSYNAFPQWS*

>G2509 (143..934)

ATATATTCCCTCTTTCATTCTCCTTCTTCGTCTTTTCTTTGTTTCTCATATTTCAAGACAT
CCTCAATTCCAAATCTTAAACCCTAAATTTACAGACACAATCGAGATCACCTGAAAAAAG
AGGTTTAAAGATTTTAGCAAAGATGGCGAATTCAGGAAATTATGGAAAGAGGCCCTTTCG
AGGCGATGAATCGGATGAAAAGAAAGAACCGATGATGATGAGAACATATTCCTTTCTT
CTCTGCCCCGATCCCAATATGACATGCGTGCCATGGTCTCAGCCTTGACTCAAGTCATTGG
AAACCAAAGCAGCTCTCATGATAATAACCAACATCAACCTGTTGTGTATAATCAACAAGA
TCCTAACCACCGGCTCCTCCAACCTCAAGATCAAGGGCTATTGAGGAAGAGGCATATAG
AGGGGTAAAGACAACGACCATGGGGAAGTGGGCAGCTGAAATTCGGGATCCGCAAAAGGC
AGCACGGGTGTGGCTCGGGACATTTGAGACTGCTGAAGCTGCGGCTTTAGCTTATGATAA
CGCAGCTCTTAAGTTCAAAGGAAGCAAAGCCAACTCAATTTCCCTGAGAGAGCTCAACT
AGCAAGTAACACTAGTACAACCTACCGGTCCACCAAACCTATTATTTCTTAATAATCAAAT
TTACTACTCAAATCCGAGACTAATCCGCAAACCATACCTTATTTTAACCAATACTACTA
TAACCAATATCTTTCATCAAGGGGGGAATAGTAACGATGCATTAAAGTTATAGCTTGGCCGG
TGGAGAAACCGGAGGCTCAATGTATAATCATCAGACGTTATCTACTACAAATTTCTTCATC
TTCTGGTGGATCTTCAAGGCAACAAGATGATGAACAAGATTACGCCAGATATTTGCGTTT
TGGGGATTCTTCACCTCTAATTCTGTTTTTGGAGATCTTCAATAAACTGATAATAAAGG
ATTTGGGTCACTTGTTATGAGGGGATCATATGTTTTCTAA

>G2509 Amino Acid Sequence (domain in aa coordinates: 89-156)

MANSGNYGKRPFRGDESDEKKEADDDENIFPFFSARSQYDMRAMVSALTQVIGNQSSSHD
NNQHQPVVYNQQDPNPPAPPTQDQGLLRKRYRGVRQRPWGKWAIEIRD PQKAARVWLGT
FETAEEAALAYDNAALKFKGSKAKLNFPERAQLASNTSTTTGPPNYSSNNQIYYSNPQT
NPQTIPIFYFNQYYNQLYHQGNSNDALSYSLAGGETGGS MYNHQTLSTTNSSSSGGSSRQ
QDDEQDYARYLRFGDSSPPNSGF*

>G390 (1..2526)

ATGATGGCTCATCACTCCATGGACGATAGAGACTCTCCTGATAAAGGATTTGATTCCGGC
AAGTACGTTAGATACAGCCGGAACAAGTTGAAGCTCTTGAGAGAGTTTATGCTGAGTGT
CCTAAACCTAGCTCTCTGAGAAGACAACAGCTTATTCGTGAATGTCCCATTCTCTGTAAC
ATCGAGCCTCGACAGATCAAAGTTTGGTTCCAGAATCGCAGATGTGAGAGAAGCAGAGG
AAAGAGTCAGCTCGTCTTCAGACAGTGAACAGGAAGCTGAGTGCTATGAACAAGCTTTTG
ATGGAAGAGAATGATCGTTTGCAGAAGCAAGTCTCCAACCTTGGTTTATGAGAATGGATT
ATGAAACATCGAATCCACACTGCTTCTGGGACGACCACAGACAACAGCTGTGAGTCTGTG
GTCGTGAGTGGTCAGCAACGTGAGCAGCAAAACCAACACATCAGCATCCTCAGCGTGAT
GTTAACAACCCAGCTAATCTTCTCTCGATTGCGGAGGAGACCTTGGCGGAGTTCCTTTGC
AAGGCTACAGGAACCTGTGCGACTGGGTCCAGATGATTGGGATGAAGCCTGGTCCGGAT
TCTATTGGTATCGTAGCTGTTTACGCAACTGCAGTGAATAGCAGCACGTCCTGTGGC
CTCGTGAGTTTAGAACCCATGAAGGTGCGCTGAAATCCTCAAAGATCGTCCATCTTGGTTC
CGTGACTGTGATGTGTGAGACTCTGAATGTTATACCCACTGGAAATGGTGGTACTATC
GAGCTTGTCAACACTCAGATTTATGCTCCTACAACATTAGCAGCAGCTCGTGACTTTTGG
ACGCTGAGATATAGTACAAGTCTAGAAGATGGAAGCTATGTGGTCTGTGAGAGATCACTC
ACTTCTGCAACTGGTGGCCCCAATGGTCCACTTTCTTCAAGCTTCGTGAGAGCCAAAATG
CTGTCAAGCGGTTTTCTTATCCGTCCTTGTGATGGTGGTGGTTCATTATTCACATCGTT
GATCATGTGGACTTGGATGTCTCAAGTGTTCCTGAAGTCCTCAGGCCTCTTTATGAGTCT
TCCAAAATCCTTGCTCAAAAATGACTGTGCTGCTCTGAGACATGTGCGCCAAATTGCT
CAAGAGACTAGTGGAGAAGTCCAGTATAGTGGTGGACGCCAGCCTGCAGTTTTTAAGGACT
TTCAGCCAGAGACTCTGCCGGGGTTTCAATGATGCTGTAAATGGTTTTGTGCGATGATGGA
TGGTCTCCAATGAGTAGTGATGGAGGAGAGGATATTACGATCATGATTAACTTCTCTCT
GCTAAATTTGCTGGCTCCTCAATACGGTAGCTCATTTCTTCCAAGTTTTGGAAGTGGTGT
CTCTGTGCCAAAGCTTCTATGCTGTTGCAGAATGTTCCACCCCTTGTATTGATTCCGTT
CTGAGAGAACACCGAGCTGAATGGGCAGACTATGGTGTGCGATGCTTCTGTGCTCATCT
CTCAGAGCAACTCCATATGCTGTTCCATGCGTCAGAACCGGTGGGTTCCCGAGTAACCAA
GTCATTCTTCTCTCGCACAGACACTCGAACATGAAGAGTTTCTCGAAGTGGTTAGACTT

GGAGGTCATGCTTACTCACCTGAAGACATGGGCTTATCCCGGGATATGTATTTACTGCAG
CTTTGTAGCGGCGTTGATGAAAAATGTGGTTGGAGGTTGTGCTCAGCTTGTCTTTGCCCA
ATCGATGAATCATTGTGCTGATGACACCTTTGCTTCCTTCTGGTTTCCGTGTCATACCA
CTCGACCAAAAAACAAATCCGAATGATCATCAATCTGCAAGTCGAACACGGGATCTAGCA
TCGTCCTTAGATGGTTCCACCAAAACCGATTCCGAAACAACTCTAGATTGGTCTTAACA
ATAGCCTTCCAGTTCACGTTTGATAACCATTCAGAGACAATGTTGCTACAATGGCGAGA
CAGTATGTGAGGAACGTTGTTGGTTCGATTTCAGAGAGTGGCTCTAGCCATTACGCCTCGT
CCTGGCTCAATGCAACTTCCCACTTCCCTGAAGCTCTCACTCTTGTCCGTTGGATCACC
CGTAGTTACAGTATTCATACAGGTGCAGATCTGTTTGGAGCTGATTCTCAGTCCTGTGGA
GGAGACACATTGCTTAAGCAACTCTGGGACCATAGTGATGCCATATGTGCTGCTCCCTG
AAAATAATGCCTCACCGGTATTCACATTTGCAAACCAAGCTGGTTTAGACATGCTTGAA
ACTACACTTGTGGCACTTCAGGATATAATGCTCGACAAAACACTTGATGACTCTGGTCGT
AGAGCTCTTTGCTCCGAGTTCGCCAAGATCATGCAGCAGGGATATGCGAATCTTCCGCA
GGAATATGTGTGTCGAGCATGGGCAGACCGGTTTCGTATGAGCAAGCGACGGTGTGGAAA
GTTGTTGATGACAACGAATCAAACCACTGCTTGGCTTTTACCCTCGTTAGTTGGTCGTTT
GTTTGA

>G390 Amino Acid Sequence (domain in AA coordinates: 18-81)
MMAHHSMDRSDSPDKGFDGSKYVRYTPEQVEALERVYAECPKPSSLRQQLIRECPILCN
IEPRQIKVWFQNRRCREKQKESARLQTVNRKLSAMNKLMEENDRLQKQVSNLVYENG
MKHRIHTASGTTTNDSCSESVVSGQQRQQNPHTQHPRQDVNNPANLLSTABETLAEFLC
KATGTAVDWVQMIGMKPGPDSIGIVAVSRNCSGIAARACGLVSLPEMKVAEILKDRPSWF
RDCRCVETLNVPTGNGGTIELVNTQIYAPTTLAAARDFWTLRYSTSLDGSYVVCERSL
TSATGGPNGLSSSFVRKMLSSGFLIRPCDGGGSI IHIVDHVDLDVSSVPEVLRPLYES
SKILAQKMTVAALRHVRQIAQETSSEVQYSGGRQPAVLRTFSQRLCRGFNDVNGFVDDG
WSPMSSDGGEDITIMINSSSAKFAGSQYSSFLPSFGSGVLCASMLLQNVPLVLIRF
LREHRAEWADYGVDAISAASLRATPYAVPCVRTGGFPNSQVILPLAQTLHEEFLEVRL
GGHAYSPEDMGLSRDMLYLQLCGVDENVVGGCAQLVFAPIDESFADDAPLLPSGFRVIP
LDQKTNPNHDQASRTRDLASSLDGSTKTDSETNSRLVLTIAFQFTFDNHSRDNVATMAR
QYVRNVVSGIQRFVALAITPRPGSMQLPTSPEALTLVRWITRSYSIHTGADLFGADSQSCG
GDTLLKQLWDHSDAILCCSLKTNASPVFTFANQAGLDMLETTLVALQDIMLDKTLDDSGR
RALCSEFAKIMQQGYANLPAGICVSSMGRPVSYEQATVWKVDDNESNHCLAFTLVSWSF
V*

>G391 (1..2559)

ATGATGATGGTCCATTTCGATGAGCAGAGATATGATGAACAGAGAGTCGCCGGATAAAGGG
TTAGATTCCGGCAAGTATGTGAGGTACACGCCGGAGCAAGTGGAAGCTCTCGAGAGAGTT
TACACTGAGTGTCTTAAGCCAAGTTCTCTAAGAAGACAACAACCTCATACGTGAATGTCCG
ATTCTCTCTAACATCGAGCCTAAGCAGATCAAAGTTTGGTTTCAGAACCGCAGATGTCTGT
GAGAAGCAGAGGAAAGAAGCTGCTCGTCTTCAAACAGTGAACAGAAAACCTCAATGCCATG
AACAAACTCTTGATGGAAGAGAATGATCGTTTGCAGAAGCAAGTTTCTAACTTGGTCTAT
GAGAATGGCCACATGAAACATCAACTTCACACTGCTTCTGGGACGACCACAGACAACAGC
TGTGAGTCTGTGGTCTGTGAGTGGTCAGCAACATCAACAGCAAAACCCAAATCCTCAGCAT
CAGCAACGTGATGTAAACAACCCAGCAGGACTCCTTTCTATAGCAGAGGAGGCCCTAGCA
GAGTTCCTTTCCAAGGCTACAGGAACTGCTGTTGACTGGGTTTCAGATGATTGGGATGAAG
CCTGGTCCGGATTCTATTGGCATACTCGCTATTTCCGCGCAACTGCAGCGGAATTGCAGCA
CGTGCTGCGGCCTCGTGAGTTTGAACCCATGAAGGTTGCTGAAATTCTCAAAGATCGT
CCATCTTGGCTCCGAGATTGTGCAAGTGTGGATACTCTGAGTGTGATACCTGCTGGAAAC
GGTGGGACGATCGAGCTTATTTACACGCAGATGTATGCTCTTACGACTTTAGCAGCAGCT
CGTGACTTTTGGACGCTGAGATATAGCACATGTTTGAAGATGGAAGCTATGTGGTTTGT
GAAAGGTGCGTTACTTCTGCAACTGGTGGCCCCACTGGGCCACCTTCTTCAAACCTTGTG
AGAGCTGAAATGAAACCAAGCGGGTTTCTCATCCGTCCTTGCGATGGTGGTGGTTCCATT
CTCCACATTGTTGATCATGTTGATCTGGATGCCTGGAGTGTCCCTGAAGTCATGAGCCCT
CTCTATGAATCATCGAAGATTCTTGCTCAGAAAATGACTGTTGCTGCTTTGAGACATGTA
AGACAAATGACAAAGAAACAAGTGGAAGTTTCAATGATGGTGGAGGGCGCAACCTGCG
GTTTTAAGAACCTTCAGTCAAAGACTCTGTCGGGGTTTCAATGATGCTGTTAATGGTTTT
GTGGATGATGATGGTACCAATGGGTAGCGATGGTGCAGAGGATGTTACTGTAATGATA
AACTTGTCCCTGGGAAGTTTGGTGGTCTCAGTACGGTAATTCAATTCCTTCCAAGCTTT
GGTAGTGGCGTGCTTTGTGCCAAGGCATCTATGTTGCTTCAGAAGTTCCACCCGCTGTG

CTGGTTCGATTCCCTTAGAGAACACCGATCTGAATGGGCTGATTATGGCGTGGATGCTTAT
GCTGCTGCATCGCTCAGAGCAAGTCCTTTTGTGCTTCTTGTGCTAGAGCTGGGGGGTTC
CCAAGTAACCAAGTCATTCTTCTCTTGGCGAGACAGTTGAACATGAAGAGTCACTTGAG
GTGGTTAGACTTGAAGGTCACGCTTACTCACCCGAAGACATGGGTTAGCTCGGGATATG
TATTTGCTACAGCTTTGTAGCGGTGTGTGATGAAAATGTGGTTGGAGGTGTGTCACAGCTT
GTATTTGCCCCATCGATGAATCATTTGCTGATGATGCACCTTTGCTTCCTTCCGGTTTC
CGCATACACTCTTGAACAGAAATCTACTCCGAACGGTGCATCTGCAAACCGTACCTTG
GATTTAGCCTCAGCTTTAGAAGGATCCACACGTCAAGCTGGTGAAGCCGACCCAAATGGC
TGTAACCTTAGGTGCGTACTAACCATAGCATTCCAGTTCACATTTGATAACCATTCAAGA
GACAGTGTGCTTCAATGGCAGCTCAGTACGTGCGAAGCATAGTAGGATCGATTAGAGG
GTTGCTCTAGCCATTGCTCCTCGTCCGCTCCAATATCAGTCCAATATCTGTTCCCACT
TCCCCTGAAGCTCTCACTCTGGTCCGTGGATCTCCCGGAGTTACAGCCTTCACACTGGT
GCAGATCTCTTTGGATCTGATTCTCAAACAGTGGTGACAGCTTGCTGCATCAACTCTGG
AATCACTCTGATGCAATCTTGTGCTGCTCCCTCAAACAAACGCTTCACCGGTTTTTACA
TTCGCAAACCAACCGGTTTAGACATGCTGGAACGACTCTTGTAGCCCTTCAAGACATA
ATGCTAGACAAGACCTTGACGAACCTGGTCGTAAAGCTCTTGTCTGAGTTCCCAAG
ATCATGCAACAGGGCTATGCTCATCTGCCGCGAGGAGTATGTGCGTCAAGCATGGGAAGG
ATGGTATCTTACGAGCAGGCAACGGTGTGGAAGTTCTTGAAGACGATGAATCAAACCAC
TGCTTAGCTTTCATGTTTCGTGAATTGGTCTGTTCTGTTGA

>G391 Amino Acid Sequence (domain in AA coordinates: 25-85)
MMMVHMSRDMNRESPPDKGLDSGKYVRYTPEQVEALERVYTECPKPSLRQQLIRECP
ILSNIEPKQIKVWFQNRRCREKQKBAARLQTVNRKLNAMNKLMEENDRLQKQVSNLVY
ENGHMKHQLHTASGTTTNDNSCESVVVSGQHQHQQNPQHQQRDANNPAGLLSIAEEALA
EFLSKATGTAVDWVQMIGMKPGPDSIGIVAISRNCSGIAARACGLVSLPEMKVAEILKDR
PSWLRCRSDVTLVSVIPAGNGGTIELIYTQMYAPTLLAAARDFWTLRYSTCLEDSYVVC
ERSLTSATGGPTGPPSSNFVRAEMKPSGFLIRPCDGGGSILHIVDHVDLDAWSVPEVMRP
LYESSKILAQKMTVAALRHVRQIAQETSSEVQYGGGRQPAVLRTFSQRLCRGFNDVNGF
VDDGWSPPMGSDGAEDVTVMINLSPGKFGGSQYGNFSLPSFGSVLCAKASMLLQNVPPAV
LVRLFLREHRSEWADYGVDAAYAAASLRASPFVPCARAGGFPSNQVILPLAQTVHEESLE
VVRLEGHAYSPEMDGLARMYLLQLCSGVDENVVGGCAQLVFAPIDESFADDAPLLPSGF
RIIPLEQKSTPNGASANRTLDLASALEGSTRQAGEADPNCGNFRSVLTIAFQFTFDNHSR
DSVASMARQYVRSIVGSIQRVALAIAPRPGSNISPISVPTSPEALTLVRWISRSYSLHTG
ADLFGSDSQTSGDTLLHLWNHSDAILCCSLKTNASPVFTFANQTGLDMLETTLVALQDI
MLDKTLDEPGRKALCSEFPKIMQQGYAHLPAVCASSMGRMVSYEQATVWKVLEDDSNH
CLAFMFVNWSFV*

>G438 (188..2716)

CGGGGTACCAAGCCACGACCGTAGAATCTTCTTTGTCTGAAAAGAATTACAATTTACG
TTTCTCTTACGATACGACGGACTTTCCGAAGAAATTAATTTAAAGAGAAAAGAAGAA
GCCAAAGAAGAAGAAGTAGCTAGAAGAAACAGTAAAGTTTGAGACTTTTTTTGAGGCTCG
AGCTAAATGGAGATGGCGGTGGCTAACCAACCGTGAGAGAAGCAGTGACAGTATGAATAG
ACATTTAGATAGTAGCGGTAAGTACGTTAGGTACACAGCTGAGCAAGTCGAGGCTCTTGA
GCGTGTCTACGCTGAGTGTCTTAAGCCTAGCTCTCTCCGTCGACAACAATTGATCCGTGA
ATGTTCCATTTTGGCCAATATTGAGCCTAAGCAGATCAAAGTCTGGTTTCAGAACCGCAG
GTGTCGAGATAAGCAGAGGAAAGAGGCGTCGAGGCTCCAGAGCGTAAACCGGAAGCTCTC
TGCGATGAATAAACTGTTGATGGAGGAGAATGATAGGTTGCAGAAGCAGGTTTCTCAGCT
TGTCTCGGAAATGGATATATGAAACAGCAGCTAACTACTGTTGTTAACGATCCAAGCTG
TGAATCTGTGGTCACAACTCCTCAGCATTCGCTTAGAGATGCGAATAGTCTGCTGGATT
GCTCTCAATCGCAGAGGAGACTTTGGCAGAGTTCTTATCCAAGGCTACAGGAACTGCTGT
TGATTGGGTTTCAGATGCCTGGGATGAAGCCTGGTCCGGATTTCGGTTGGCATCTTTGCCAT
TTCGCAAAGATGCAATGGAGTGGCAGCTCGAGCCTGTGGTCTTGTGTAGCTTAGAACCTAT
GAAGATTGCGAGATCCTCAAAGATCGGCCATCTTGGTTCCGTGACTGTAGGAGCCTTGA
AGTTTTCACTATGTTCCCGGCTGGTAATGGTGGCACAATCGAGCTTGTATATGCGAGAC
GTATGCACCAACGACTCTGGCTCCTGCCCGGATTCTTGACCCCTGAGATACACAACGAG
CCTCGACAATGGGAGTTTTGTGGTTTGTGAGAGGTCGCTATCTGGCTCTGGAGCTGGGCC
TAATGCTGCTTCAGCTTCTCAGTTTGTGAGAGCAGAAATGCTTTCTAGTGGGTATTTAAT
AAGGCCTTGTGATGGTGGTGGTCTATTATTACATTGTGATCACCTTAATCTTGAGGC
TTGGAGTGTTCGGATGTGCTTCGACCCCTTTATGAGTCATCAAAGTCGTTGCACAAAA

AATGACCATTTCCGCGTTGCGGTATATCAGGCAATTAGCCCAAGAGTCTAATGGTGAAGT
 AGTGTATGGATTAGGAAGGCAGCCTGCTGTTCTTAGAACCTTTAGCCAAAGATTAAGCAG
 GGGCTTCAATGATGCGGTTAATGGGTTTGGTGACGACGGGTGGTCTACGATGCATTGTGA
 TGGAGCGGAAGATATTATCGTTGCTATTAACCTCTACAAAGCATTTGAATAATATTTCTAA
 TTCTCTTTTCGTTCCCTTGAGGGCGTGCTCTGTGCCAAGGCTTCAATGCTTCTCCAAAATGT
 TCCTCCTGCGGTTTTTTCGTTCCCTTAGAGAGCATCGATCTGAGTGGGCTGATTTCAA
 TGTTGATGCATATTCCGCTGCTACACTTAAAGCTGGTAGCTTTGCTTATCCGGGAATGAG
 ACCAACAAGATTCCTGGGAGTCAGATCATAATGCCACTAGGACATACAATTGAACACGA
 AGAAATGCTAGAAGTTGTTAGACTGGAAGGTCATTCTCTTGCTCAAGAAGATGCATTTAT
 GTCACGGGATGTCCATCTCCTTCAGATTTGTACCGGGATTGACGAGAATGCCGTTGGAGC
 TTGTTCTGAAGTATTTGCTCCGATTAATGAGATGTTCCCGGATGATGCTCCACTTGT
 TCCCTCTGGATTCCGAGTCATACCCGTTGATGCTAAAACGGGAGATGTACAAGATCTGTT
 AACCGCTAATCACCCTACACTAGACTTAACCTTAGCCTTGAAGTCGGTCCATCACCTGA
 GAATGCTTCTGGAAACTCTTTTCTAGCTCAAGCTCGAGATGTAATCTCACTATCGCGTT
 TCAATTCCTTTTGAACAACTTGCAAGAAAATGTTGCTGGTATGGCTTGTCAGTATGT
 GAGGAGCGTGATCTCATCAGTTCAACGTGTTGCAATGGCGATCTCACCGTCTGGGATAAG
 CCCGAGTCTGGGCTCCAAATTGTCCCAGGATCTCTGAAGCTGTTACTCTTGCTCAGTG
 GATCTCTCAAAGTTACAGTCATCACTTAGGCTCGGAGTTGCTGACGATTGATTCACCTTGG
 AAGCGACGACTCGGTACTAAAACCTTATGGGATCACCAAGATGCCATCCTGTGTTGCTC
 ATTAAGCCACAGCCAGTGTTTATGTTTGGCAACCAAGCTGGTCTAGACATGCTAGAGAC
 AACACTTGTAGCTTACAGATATAACACTCGAAAAGATATTGATGAATCGGGTCGTAA
 GGCTATCTGTTGCGACTTCGCCAAGCTAATGCAACAGGGATTGCTTGCTTGCCTTCAGG
 AATCTGTGTGTCACGATGGGAAGACATGTGAGTTATGAACAAGCTGTTGCTTGGAAAGT
 GTTTGCTGCATCTGAAGAAAACAACAATCTGCATTGTCTTGCCTTCTCCTTTGTAAA
 CTGGTCTTTTGTGTGATTGATTGACAGAAAAGACTAATTTAAATTTACGTTAGAGAAC
 TCAAATTTTGGTTGTGTTTAGGTGCTCTGTTTTGTTTTTAAATTTATTTTGATCAA
 A

>G438 Amino Acid Sequence (domain in AA coordinates: 22-85)
 MEMAVANHRRSSDSMNRLDSSGKYVRYTAEQVEALERVYAECPKPSSLRRQQLIRECS
 ILANIEPKQIKVWFQNRRCRDKQRKEASRLQSVNRKLSAMNKLMEENDRLQKQVSQLVC
 ENGYMKQLTIVVNDPSCSVVTTTQHSRLDANSFAGLLSIAEETLAEFLSKATGTAVDW
 VQMPGMPKPGPDSVGIFAIQSQRNGVAARACGLVSLPEMKIAEILKDRPSWFRDCRSLEVF
 TMFPAGNGGTIELVYMQTYAPTTLAPARDFWTLRYTSLDNGSFVVCERSLSGSGAGPNA
 ASASQFVRAELSSGYLIRPCDGGGSIHIVDHLNLEAWSVPDVLRLPLYESSKVVAQKMT
 ISALRYIRQLAQESNGEVVYGLGRQPAVLRTFSQRLSRGFNDVNGFGDDGWSTMHCDGA
 EDIIVAINSTKHLNNISNLSFLGGVLCASMLLQNVPPAVLIRFLREHRSEWADFNVD
 AYSAAATLKAGSFAYPGMRPTRFTGSQIIMPLGHTIEHEEMLEVVRLEGHSLAQEDAFMSR
 DVHLLQICTGIDENAVGACSELI FAPINEMFPDDAPLVP SGFRVIPVDAKTGDVQDLLTA
 NHRTL DLTSSLEVGPSPENASGNSFSSSSSRILTIAFQFPFENNLOENVAGMACQYVRS
 VISSVQRVMAISP SGISPLSGSKLSPGSPEAVTLAQWISQSYSHHLGSELLTIDSLGSD
 DSVLKLWDHQDAILCCSLKPQPVFMFANQAGLDMLETTLVALQDITLEKIFDESGRKAI
 CSDFAKLMQGGFACLPSGICVSTMGRHVSIEQAVAWKVFAASEENNNNLHCLAFS FVNWS
 FV*

>G47 (38..472)

CTTCTTCTTCACATCGATCATCATACAACAACAAAAATGGATTACAGAGAATCCACCGG
 TGAAAGTCAGTCAAAGTACAAAGGAATCCGTCGTCGGAAATGGGGCAAATGGGTATCAGA
 GATTAGAGTTCCGGGAACCTGTCACCGTCTCTGGTTAGGTTCAATCTCAACAGCAGAAGG
 TGCCGCGGTAGCACACGACGTTGCTTTCTGTTTACACCAACCTGATTCTTTAGAATC
 TCTCAATTTCCCTCATTTCGCTTAATCCTTCACTCGTTTCCAGAACCTTCCGAGATCTAT
 CCAGCAAGCTGCTTCTAACGCCCGCATGGCCATTGACGCCGGAATCGTCCACAGTACCAG
 CGTGAACCTCTGGATGCGGAGATACGACGACGTATTACGAGAATGGAGCTGATCAAGTGGA
 GCCGTTGAATATTTAGTGATGATTATCTGGGCGGCCACGATCACGTTGATTATCTC
 GACGGTCATGATCACGTTTGATCTTCTTTGAGTAAGATTTTGTACCATAATCAAAACAG
 GTGTGGTGCTAAAATCTTACTCAAAACAAGATTAGGTACCAAGAGAAAACAATCAAATGG
 TTGTGAATATACATTATAAGGTTTTGATTAAATGTTTGTTCCTGATTAGTGAAGTTG
 GTCCATTGTATACAAATCTATTCAAGAAACCTAGCGCGAGATCATGTTTCGTGATTGAAG
 ATTGAGATTTTAAAGTATTCGTAATATTTTGTAAATACAAATAAAAAAAAAAAAAAAA

AAAAA

>G47 Amino Acid Sequence (domain in AA coordinates: 11-80)
MDYRESTGESQSKYKGI RRRKWKVSEIRVPGTRDRLWLGSFSTAEGAAVAHDVAFFCL
HQPDSLES LNFP HLLNPSLVSR TSPRSIQQAASNAGMAIDAGIVHSTSVNSGCGD TTTY
ENGADQVEPLNISVYDYLGGHDHV*

>G559 (89..1285)

aaagttgctagcttttaatttgccaacttactattcttatgtgtaataatcgtttgagg
tcggtgatttgggtgataagtcagtagaaATGgataaggagaaatctccagcacctccttg
tggaggtcttcctcctccatctccatcaggtcgatgctctgcattctcagaagctggtcc
cattggtcatggttcagatgctaatacgatgagtcgatattagccgatgcttgataa
cccacctaagaagattggacatcggcgagctcattctgaaatacttactctccctgatga
tttgagcttggatagtgatcttgggtggttggaatgctgctgatggagcttcttctc
tgatgagactgaagaagatttgctctctatgtatcttgatagggataagtttaattcttc
tgctacatcttctgcccagttggtgagccatcaggaactgcttggaataatgagacaat
gatgcagacaggcacaggctcaacttccaatcctcagaatacggtaatagtcttggcga
aaggccaagaatcaggcatcaacatagccaatctatggatggttcaatgaatatcaatga
gatgcttatgtcgggaaatgaagatgattctgctattgatgctaagaagtctatgtctgc
tactaaacttgctgagcttgctctcattgatcctaactgctgaaggatagggcaaa
caggcagtcgcgcagcagatcaaaagaaaggaagacgagatacatatttgagcttgagag
aaaagtacagacttggcaaacagaggctacaactctctcagcccagttgacctcttaca
gagagacacaaatggcttgactggtgaaaacaatgagctgaagctgcggttacaacaat
ggagcagcaggttcacttgaggatgaactaaacgaagcactaaaggaggaaatccagca
tctgaaggtgttgactggccaggttgctccatcagcgttgaactatgggtcggttggtc
aaaccagcagcaattctattccaacaatcagtcgaatgcaacaatcttagctgcaaaaca
gttccagcaacttcagattcattcacagaagcagcaacaacaacaacaacaacaaca
gcaacaccaacagcagcagcagcaacagcaacagtcagtttcaacagcaacagatgca
acagcttatgagcagcgggttcaacagcaagaacaacaaatggagtaagactcaagcc
ttcacaagcccagaaagagaacTGAggaatatgaatatgtcccacgtaagtgaaggttc
tccttctgaacaattcctttctcattcataaattggttggtcattccatcacttgagcttc
ttggatttttagggtttttagctaacaca

>G559 Amino Acid Sequence (domain in AA coordinates: 203-264)

MDKEKSPAPPCGGLPPPSPSGRCSAFSEAGPIGHGSDANRMSHDISRMLDNPPKKIGHRR
AHSEILTL PDDL SFDSDLGVVGN AADGASFSDETEEDLLSMYLDMDKFNSSATSSAQVGE
PSGTAWKNETMMQTGTGSTSNPQNTVNSLGERPRI RHQHSQSM DGS MNIN EMLMSGNEDD
SAIDAKKSMSATKLAEALALIDPKRAKRIWANRQSAARSKERKTRYIFELERKVQTLQTEA
TTL SAQL TLL QRD TNGLTVENNELKLRLQ TMEQQVHLQDELNEALKEEI QHLKVL TGQVA
PSALNYGSFSGNQQQFYNNQSMQTILAAKQFQQLQIHSQKQQQQQQQQQQHQQQQQQQ
QQYQFQQQQMQQLMQRLQQEQNGVRLKPSQAQKEN*

>G568 (141..995)

GACCGGCTAAAGTCAAGAACCTCTCTCTGAGCTCTCACCCTTTCTCTCTACTCCCTC
TCTGCGTGTAGGATACTACTAGACAATTGACAACCAAAGACTAAAGCTGTGTGTTGGTT
CACTTCTGTTCTCTTTTCCAATGTTGTCATCAGCTAAGCATCAGAGAAACCATAGACTCT
CTGTACAAACAAGAACCAGACTCTCACCAGTCTTCTTCCATTTTCATCCTCATCACCAT
CGTCTTCTTCTCATCATCATCAACCTCATCATCTCCTTTACCTTCTCAAGACTCTC
AAGCCCAGAAGAGATCTCTTGTCACCATGGAAGAAGTTTGAATGACATCAACCTTGCTT
CCATCCACCCTAAACCGACACAGCCCTCATCCACAACACAACCACGAGCCAAGGTTCA
GGGGCCAAAACCACCAACCAACCAACCTAACTCAATCTTCAAGATTTTCTCAAAGGAT
CTTTGAACCAGGAACAGCACCCACAAGCCAGACCACGGGTTCTGCGCCTAATGGCGATT
CCACCACGGTCACTGTTCTTTACAGCTCTCCTTTTCCACCTCCTGCAACTGTTCTGAGCT
TGAATTCGCGGCTGCTTTCGAGTTCTCGATAACCAAGATCCTCTTGTACCTCAAACCT
CTAATCTTCATACCCACCATCACCTCTCAAACGCTCATGCCCTCAACACCTCTTTCGAGG
CTCTGGTTCCATCCAGTTCTTTTGGTAAGAAAAGAGGCCAAGATTCCAATGAAGGTTTCA
GGAAATAGAAGACATAAGCGTATGATCAAGAACAGAGAATCTGCAGCTCGTTCCCGCGCTA
GGAAACAGGCTTATACAAACGAGTTAGAACTTGAAGTTGCTCACTTGCAGGCAGAAAATG
CAAGACTCAAGAGACAACAAGATCAAAAAATGGCTGCAGCAATTGAGCAACCCAAAAAGA
ACACACTTCAACGGTCTTCCACAGCTCCATTTTGAGAAATCTACAAGTCTTGTCTCTCT
TTTGGGGATTGAGATTGTCTCATGAAGAAGTGAAGAAAATGGCAAAAGTTTGTACCTTTT

TTATTAGCTATAAGTATAACTAAGCCTAAAATTGTAGAACTAAGATATTGTAGGGGAAAA
AAGAAGATGTAAACAAAAGACCCGGAAAGAGAAAAGGATCTTTCAATTTCTTAAGGCAC
AGGAACACCTGTCTGGGTCTCTCTTAATGTTCTGTCTGTTTTCCTATGCAAACCCCTTT
TTCACCTCTGTACTAACTTATACTTGTATTCTTG
>G568 Amino Acid Sequence (domain in AA coordinates: 215-265)
MLSSAKHQNRNHLRSATNKNQTLTKVSSISSSSPSSSSSSSSSSSSPLPSQDSQAQKRSL
VTMEEVWINDINLASIHHLNRHSPHPQHNHEPRFRGQNHNNQNPNSIFQDFLKGSLNQEP
PTSQTTGSAPNGDSTTVTVLYSSPFPPTATVLSLNSGAGFEFLDNQDPLVTSNSNLHTHH
HLSNAHAFNTSFEALVPSSSFGKKRGQDSNEGSGNRRHKRMIKNRESAARSARKQAYTN
ELELEVAHLQAENARLKRQDQKMAAAIQPKKNTLQRSSTAPF*
>G580 (43..747)
CCAAAAACAAAGCATTCTATGCTATTCTGTTCTGTTCTCCAATGTTGTCATCAGCAAAG
CATAATAAGATCAACAACCATAGTGCCTTTTCAATTTCTCTTCATCATCATCATTTATCA
ACATCATCTCCCTAGGCCATAACAAATCTCAAGTCACCATGGAAGAAGTATGGAAAGAA
ATCAACCTTGGTTCACCTTCACTACCATCGGCAACTAAACATTGGTCATGAACCAATGTTA
AAGAACCAAAACCTTAATAACTCCATCTTTCAAGATTTCTCAACATGCCTCTGAATCAA
CCACCACCACCACCACCACCCTTCTCTTCCACCATTGTCACTGCTCTCTATGGCTCT
CTGCCTCTTCCGCTCTCTGCCACTGTCTCAGCTTAAACTCCGGTGTGGATTGAGTTT
CTTGATACCACAGAAAATCTTCTGCTTCTAACCCTCGCTCCTTTGAGGAATCTGCAAAG
TTTGGTTGTCTTGGTAAGAAAAGAGGCCAAGATTCTGATGATACTAGAGGAGACAGAAGG
TATAAGCGTATGATCAAGAACAGAGAATCTGCTGCTCGTTCAAGGGCTAGGAAGCAGGCA
TATACAAACGAACCTTGAGCTTGAAATTGCTCACTTGCAGACAGAGAATGCAAGACTCAAG
ATACAAACAGAGCAGCTGAAAATAGCCGAAGCAACTCAAAACCAAGTAAAGAAAACACTA
CAACGGTCTTCCACAGCTCCATTTTGAGAAAATCTACTATTTCTTTTGGGGGAGTTTC
AAGTGTCTTCTATGAAGATGAGAAAAACAGAAAAGTTTGTACATTTTAGCTAAGTTAAA
TTTGTGGTGGTAAGTAATGTAAAAGAAAAGTGTGTGTAGAAGAAAAGTGTCTAGAAAAAG
AAAGCAACTAATCTTCTTCTTCTCTCTGTTTCTCTATCAACTCTTTTGACTTTTGTACT
TTTTTCTTCTCTACTTAACTCTATTATTGTAATGCCAAGTCAAGTCCTTATCTAGCTA
GTACATGAGTTTCTGTTTTCACTGGTTAAGCCAT
>G580 Amino Acid Sequence (domain in AA coordinates:162-218)
MLSSAKHNKINNHSFAFSSSSSSSLSTSSSLGHNKSQVTMEEVWKEINLGLSLHYHRQLNI
GHEPMLKNQNPNSIFQDFLNMPLNQPPPPPPSSSTIVTALYGSPLPLPPATVLSLNS
GVGFEFLDTTENLLASNPRSFEBESAKFGCLGKKRGQDSDDTRGDRRYKRMKNRESAARS
RARKQAYTNELELEIAHLQTENARLKIQQEQLKIAEATQNVKKTQRSSTAPF*
>G615 (197..1252)
TTTTTCTTTTCTTCTTTTTTTTTGCTGGTGTGAGAAATTGTACGCTTACTATCTCTCTCT
CTCTCTGCCAGATTCTCTTTTTTGTATGATGTGAAAGTTGTGCTTTTGTTCCTTAAGAAA
AAGGCATATTTTAACTTGTATCTTGGTCTTGTATCTTGATTCTTGGTTTTTTTTTAG
CTTCTTAAGTTCGGTGATGTCTCTTCCACCAATGACTACAACGATGGTAATAACAATGG
AGTGTAACCTCTCTCTTTTACCTTTCTTCACTCTTGGCCATCAAGACATCATTCATAA
TCCCTACAACCATCAGTTAAAGCATCTCCGGGCCATATGGTATCAGCAGTTCCTGAATC
TCTGATCGATTACATGGCGTTTAAGTCAAATAATGTTGTGAATCAACAAGGCTTTGAGTT
TCCTGAGGTGTCAAAGGAAATCAAGAAGGTGGTGAAGAAGGACCGACATAGCAAGATTCA
AACGGCAACAAGGGATAGAGACAGGAGGGTTAGGCTTTTTATTGGGATTGCTCGCCAATT
CTTTGATCTTCAGGATATGTTGGGGTTTGATAAAGCTAGTAAACGTTAGACTGGCTGCT
CAAGAAGTCAAGAAAAGCCATCAAGAGGTGCTACAAGCAAAAACCTCAACAATGATGA
TGAAGATTTTGGAAACATTGGAGGCGATGTAGAACAAGAAGAGGAGAAGGAGGAGGATGA
CAATGGCGATAAGAGCTTCGTGTATGTTTGAGCCCCGGGTACGTTGAAGAAGAAGTGGT
ATGTGAGGCCACGAAGGCAGGGATAAGAAAGAAGAAGAGTGAGTTGAGAAACATCTCATC
AAAGGGGCTAGGAGCCAAAGCTAGAGGAAAAGCAAAGGAGCGAACAAAAGAGATGATGGC
CTATGATAATCCAGAGACTGCCTCTGATATTACACAATCTGAAATCATGGACCCATTCAA
GAGGTCTATAGTCTTCAATGAAGGAGAAGATATGACACACCTTTTCTACAAGGAACCAAT
CGAGGAGTTTGATAATCAAGAATCTATCTTAACCAATATGACTCTACCAACGAAGATGGG
TCAAAGTTACAATCAAAATAATGGGATACTTATGTTGGTAGATCAGAGTTCTAGCAGCAA
CTATAATACATTTCTGCCTCAAAATTTGGATTATAGTTATGATCAAAACCTTTTTCATGA
CCAAACCTTATATGTAGTCAACGACAAAAATTTCCCAAGGTTTCTATAAATCTCGAC
AGTTTTGAAGGACTATGCATGATCAAGTTAAACATGTAAGCCAATATAGTCCCTTATTC

CTCTGAATGTATACAAAATCTATAGTTATGTATATCTGTTCCCTTTTAAACGTATCTTTAT
TGATCTTCTGTGCTTGATCAAAATTGTCTATTTAAGATTCAAGTTTGTGTAATATTTAG
CTACAACTTTTAAAGTGGTATTATTGTAACCTTTTGAACATATATTTTGAAGATGAATAA
GAACATGTTTATATAAAAA

>G615 Amino Acid Sequence (domain in AA coordinates:88-147)
MSSSTNDYNDGNMNGVYPLSLYLSSLSGHQDIIHNPYNHQLKASPGHMVSAVPESLIDYM
AFKSNVNVNQGFPEFPEVSKEIKKVVKDRHSKIQTAGGIRDRVRLFIGIARQFFDLQD
MLGFDKASKTLDWLLKKSRAIKEVVQAKNLNNDDEDFGNIGGDVEQEEKEEDDNGDKS
FVYGLSPGYGEEVVEATKAGIRKKKSELNRISSKGLGAKARGKAKERTKEMMAYDNPE
TASDITQSEIMDPFKRSIVFNEGEDMTHLFYKEPIEEFDNQESILTNMTLPTKMGQSYNQ
NNGILMLVDQSSSSNYNTFLPQNLDSYDQNPFFHDQTLVYVTDKNFPKGFL*

>G732 (73..588)
AAAAAAACCAAAACATAAAACTCTGTCCTTTTTTGTCTTCTTGTAACTTTTCT
TGTTAAAAATCAATGGCGTCATCTAGCAGCACATACCGGAGCTCAAGCTCTCCGACGGT
GGTAATAATAACCCGTCGGACTCCGTCGTCACCGTCGACGAAACGAAAACGTAAGAATG
TTATCGAACAGAGAATCTGCACGTAGGTCAAGGATGCGTAAACAGAAACGTTGATGAT
CTAACGGCTCAGATCAATCAGCTATCAAACGACAACCGTCAGATCTTGAACAGCCTCACC
GTAACATCTCAGCTTTACATGAAGATCCAAGCCGAGAACTCTGTTCTCACCGCTCAGATG
GAGGAGCTTAGCACCAGACTCCAATCTCTCAACGAGATCGTTGATCTTGTTCATCCAAC
GGTGCAGGATTTGGTGTGACCAGATCGACGGCTGTGGTTTTGATGATCGTACGGTTGGG
ATCGACGGATATTACGATGATATGAATATGATGAGTAATGTTAATCATTGGGGTGGTTCG
GTTTACACTAACCAACCCATTATGGCTAATGATATCAATATGTATTGATTAATAAAATTA
ATTAAATAAATTAGATGCCCTTTTTTGTCTTTTATTTTAAATTTAGCCCATTTGGT
GTTTTTGGGTTGGTGTGATGATGTAATTATAGTACATGCATCTTTGATTGGTTGGAAGGA
TAAATATAAACTTTATATATATATTGGGGCATATATATATGAGTTGTACTTTGCATGTAT
TGGTGTGTGTTTTGTTATAATTATATGATTATATATGTTTATGTTAAAAAAA

>G732 Amino Acid Sequence (domain in AA coordinates: 31-91)
MASSSSTYRSSSSSDGNNNPSDSVTVDERKRKRLSNRESARRSRMRKQKHVDDLTAQ
INQLSNDNRQILNSLTVTSQLYMKIQAENSVLTAQMEELSTRQLSLNEIVDLVQSNAGF
GVDQIDGCGFDDRTVGIDGYDDMMMSNVNHWGGSVYTNQPIMANDINMY*

>G988 (1..1338)
ATGCTTACTTCCCTTCAAATCCTCTAGCTCCTCCTCCGAAGATGCCACCGCTACCAACCAC
GAGAATCCTCCTCCTTTGTGCATCGCCTCCTCCTCGGCCGCAACCTCCGCCTCACATCAC
CTCCGTCGTCCTTCTTTTACCCTGCGAATTCGTCTCCAGTCAAACCTTCCGCGCT
CAAACCTTACTCTCAATCCTCTCCCTTAACTCTTCTCCTCACGGCGACTCCACCGAGCGA
CTTGATACCTCTTCACTAAAGCCTTGTCCGTACGAATCAACCGTCAGCAACAAGATCAG
ACGGCTGAAACGGTTGCCACGTGGACGACGAAACGAAATGACGATGAGTAACCTCCACGGTG
TTCACGAGCAGTGTATGCAAGAAGCAGTTCTTGTTCGAACCAAGAACAACCAATTCTGAC
TTCGAGTCTTGTACTATCTTTGGCTAAACCAACTAACGCCGTTTATTGCGTTTCGGTCAT
TTAACGGCGAACCAAGCTATCCTCGACGCGACGGAGACAAACGATAACGGAGCTCTACAT
ATACTTGATTTAGATATATCACAAGGACTTCAATGGCCTCCATTGATGCAAGCCCTAGCA
GAGAGGTCACTAAACCCCTAGCAGTCCACCTCCATCTCTCCGCATAACCGGATGCGGTCGA
GATGTAACCGGATTAAACCGAACTGGAGACCGGTTAACCCGTTTCGCTGACTCTTTAGGT
CTCCAATTCAGTTTACACGCTAGTGATCGTAGAAGAAGATCTCGCCGGACTTTTGCTA
CAGATCCGATTGTTAGCTCTCTCAGCCGTACAAGGAGAGACCATTGCCGTCAATTGTGTT
CACTTCCTCCACAAAATATTTAACGACGATGGAGATATGATCGGTCACTTCTTGTACGG
ATCAAGAGCTTAACTCTAGAATCGTTACAATGGCAGAGAGAGAAGCTAATCATGGAGAT
CACTCGTTCTTGAATAGATTCTCTGAGGCAGTGGATCATTACATGGCGATCTTTGATTCTG
TTGGAAGCGACGTTGCCGCCAAATAGCCGAGAGAGACTAACCCTAGAGCAACGGTGGTTC
GGTAAGGAGATTTTGGATGTTGTGGCGCGGAAGAGACGGAGAGAAAGCAAAGACATCGG
AGGTTTGAGATTTGGAAGAGATGATGAAGAGGTTTGGTTTCGTTAACGTTCTTATTGGA
AGCTTTGCTTTGTCTCAAGCTAAGCTTCTTCTTAGACTTCATTATCCTTCAGAAGGTTAT
AATCTTCAGTTCCCTTAACAATTCTTGTCTTGGCTGGCAAAATCGTCCCTCTCTCTC
GTTTCGTCGTGGAATGA

>G988 Amino Acid Sequence (domain in AA coordinates:178-195)
MLTSFKSSSSSEDAATTTTENPPPLCIASSSAATSASHHLRRLFTAANFVSQSNFTAA
QNLLSILSLNSSPHGDSTERLVHLFTKALSVRINRQQDQTAETVATWTTNEMTMSNSTV

FTSSVCKEQFLFRITKNNSDFESCYYLWLNQLTPFIRFGHLTANQAILDATETNDNGALH
ILDLDISQGLQWPPLMQALAERSSNPSSPPPSLRITGCGRDVTGLNRTGDRLTRFADSLG
LQFQFHTLVIVEEDLAGLLQLIRLLALSAVQGETIAVNCVHFLHKIFNDGDMIGHFLSA
IKSLNSRIVTMAEREANHGDSFLNRFSEAVDHYMAIFDSLEATLPPNSRERLTLEQRWF
GKEILDVVAEETERKQRHRRFEIWEEMMKRFGFVNVPISFALSQAKLLRLHYPSEGY
NLQFLNLSFLGWQNRPLFSVSSWK*

>G1519 (1..1146)

ATGAGGCTTAATGGGGATTCCGGTCCGGGTCAGGATGAACCCGGTTCGAGCGGGTTTCAC
GGCGGAATCAGACGATTCCCGTTAGCAGCTCAGCCGAGATTATGAGAGCTGCTGAGAAA
GACGATCAATACGCTTCTTTTCATCCACGAAGCTTGCCGCGATGCCTTCCGACACCTTTTC
GGTACAAGAATCGCTCTTGCTTACCAGAAGGAGATGAAGCTACTTGGACAGATGCTTTAC
TATGTTCTTACGACAGGTTTCAGGGCAACAACTTTAGGAGAGGAATATTGTGACATTATA
CAGGTTGACAGGGCCTTATGGACTCTCTCTTACACCAGCTAGACGTGCTTTGTTCATATTG
TACCAGACCGCAGTTCCATATATCGCAGAGAGAATTAGCACTCGAGCTGCTACGCAAGCA
GTCACCTTTGATGAGTCTGATGAGTTTTTGGTGATAGTCAATCCACTCACCAGAAGATG
ATAGATCTTCCATCTTCACTCAAGTTGAACTTCACTTCTGTAGTATCTAGGTTAAAC
GATAGACTTATGAGATCGTGGCACCAGCTATTCAGCGATGGCCTGTGGTTCTTCTGTGTT
GCCCGGAAGTCTTACAACCTGGTTTTGCGTGCCAATCTGATGCTTCTTACTTTGAAGGT
TTTTATTATCATATATCGAAACGTGCATCCGGGGTTCGTTATGTTTTCATAGGAAAGCAA
CTGAATCAGAGACCTAGATACCAAACTTCTGGGGTTTTCTTCTAATCCAATTGTGCATC
CTTGCTGCTGAGGGCTTGCGTCCGAGTAATTTGTCTATCTACTAGCTCCATTTCAGCAG
GCTTCTATAGGATCTTATCAAACTTCAGGAGGGAGAGGTTTACCTGTTTTAAATGAAGAG
GGGAATTTGATAAATTCGGAAGCTGAAAAGGGAAGTGGTCTACCTCCGATTCAACTTCA
ACGAGGCGAGTAGGGAATGCACCTCTCTGCTTAAGCACCCGTCAGCACCCAACGGCCACT
CCTTGTTGTCATGTGTTTTGTTGGAGCTGCATTATGGAATGGTGCAACGAGAAGCAAGAA
TGCCCTCTTTGTGCAACGCCCAATACCCATTCAAGTTTGGTTTGTGTTATCATTCTGAT
TTTTAG

>G1519 Amino Acid Sequence (domain in AA coordinates: 327-364)

MRLNGDSGPGQDEPGSSGFHGGIRRFPLAAQPEIMRAAEKDDQYASFIHEACRDAFRHLF
GTRIALAYQKEMKLLGQMLYYVLTGSGQQTLEGEYCDIIQVAGPYGLSPTPARRALFIL
YQTAVPYIAERISTRATQAVTFDESDEFFGDSHIHSPRMLDLPSSQVETSTSVVSRIN
DRLMRSWHRAIQRPVPLPVAREVLQVLRLANLMLFYFEGFYHISKRASGVRYVFIGKQ
LNQRPYQILGVFLLIQLCILAEGLRRSNLSSITSSIQQASIGSYQTSGGRGLPVLNEE
GNLITSEAEKGNWSTSDSTSEAVGKCTLCSTRQHPTATPCGHVFCWSCIMEWCNEKQE
CPLCRTPNTHSSLVCLYHSDF*

>G374 (1..1359)

ATGGACAACAAAAATGATCAGGATATTGATGTTAGATCAGTGGTTGAAGCTGTTTCCGCC
GATCTTTCTTTGGTGCTCCCTCTATGTGGTTGAGAGCATGTGCATGCGCTGCCAAGAA
AATGGAACAACAGATTTCTATTGACCTTAATTCCTCACTTCAGAAAGGTCTTAATATCT
GCATTTGAATGTCCGCATTGCGGGGAAAGGAATAATGAAGTTCAAGTTTCGAGGCGAGATT
CAACCCCGTGGATGCTGTTACAATCTAGAGGTTCTAGCTGGTGATGTGAAGATATTTGAC
CGGCAAGTTGTGAAATCTGAATCAGCCACTATTAAGATTCTTGAAGTGGATTTTGAGATT
CCACCAGAGGCCCAACGTGGAAGTTTGTCTACTGTGGAAGGGATATTAGCACGGGCTGCT
GATGAAGTGAAGTCCCTTCAAGAAGAAGCAAGAAAGTTGATCTTAAACTGCTGAAGCA
ATAGACCAATTCTGTCCAACTGAGAGCTTGTGCTAAAGCAGAGACATCCTTCACTTTC
ATTTTGGATGATCCTGCTGGAAACAGTTTTCATTGAGAACCACATGCTCCATCACCAGAT
CCCTCTCTAACCATCAAATTTATGAGCGAACACCAGAGCAACAAGCAACACTTGGATAT
GTTGCTAACCCTCTCAGGCTGGACAATCAGAAGGAAGCCTTGGCGCACCTGTGATGACT
TTCCCTTCACTTGGCGAGCATGTACGGAGCCGTGTGAGACACGGATGTTCAAATAGAA
ATCCCGTACTTTCAGGAAGTTATTGTCTATGGCATCTACATGTGACAGTTGTGGCTATCGT
AATTCTGAGTTGAAGCCTGTGGTGCAATTCCTGAAAAGGGAAAGAAGATTACTCTCTCT
GTGAGGAACATTACAGACCTTAGCCGAGATGTTATCAAGTCGGACACTGCAGGAGTGATA
ATCCCGAAGTTGATCTGGAGCTAGCTGGTGGTACACTTGGTGGAATGGTAACAACAGTT
GAAGGGTTGGTTACACAGATCAGAGAAAGCCTAGCGAGAGTTACGGATTCACTTTTGGT
GATAGTATGGAAGAGAGTAAGTTGAACAAATGGAGAGAATTTGGAGCCAGGCTCACTAAG
CTCCTAAGCTTTGAACAGCCGTGGACATTGATTCTTGATGATGAATTAGCAAATTCCTTT
ATTGCACCAGTAACAGATGATATCAAAGATGACCATCAGCTCACATTTGAAGAGTACGAG

14

MAGFDENVAVMGEWVPRSPSPGTLFSSAIGEEKSSKRVLERELSLNHGQVIGLEEDTSSN
HNKSSQSNVFRGGLSERIAARAGFNAPRLNTEINRTNTDFSIDSNLRSPCLTISSPGLS
PATLLESPVFLSNPLAQPSPTTGKFPFLPGVNGNALSSEKAKDEFFDDIGASFSPHVS
SSSSFFQGTTEMMSVDYGNYNRRSSSHQSAEEVKPGSENIENLYGIETDNQNGQNKTS
DVTNTSLETVDHQEEEEQRRGDSMAGGAPAEADGYNWRKYGQKLVKGSEYPRSYKCTN
PNCQVKKKVERSREGHITETIIYKGAHNHLKPPNRRSGMQVDGTEQVEQQQQQRDSAATW
VSCNNTQQQGGSNENNVEEGSTRFEYGNQSGSIQAQTGGQYESGDPVVVDASSTFSNDE
DEDDRGTGHSVSLGYDGGGGGGGGEGDESESKRRKLEAFAAEMSGSTRAIREPRVVVQTT
SDVDILDDGYRWRKYGQKVVKGPNPRSYKCTAPGCTVRKHVERASHDLKSVITTYEGK
HNHNDVPAARNSSHHGGGDSGNGNSGGSAAVSHHYHNGHHSEPPRGRFDRQVTTNNQSPFS
RPFSPQPHLPPSPGFSFGLGQTGLVNLSPGLAYGQKMPGLPHPYMTQPVGMSEAMMQR
GMEPKVEPVSDSGQSVYNQIMSRLPQI*

>G1000 (1..954)

ATGGGAAGACCTCCTTGTTGTGACAAGTCCAATGTCAAGAAAGGTCTCTGGACCGAGGAA
GAAGACGCTAAGATCCTTGCTTATGTTGCTATCCATGGTGTAGGAACTGGAGCTTGATC
CCCAAAAAGCAGGTCTGAATCGATGTGAAAGAGCTGTAGACTAAGATGGACTAATTAC
TTAAGACCTGACCTTAAACATGACAGCTTCTTACCCAAGAAGAAGAGCTTATCATTGAG
TGTCATAGAGCCATTGGCAGCAGGTGGTCTTCCATTGCACGAAAGCTTCCAGGAAGAACG
GATAATGATGTGAAGAATCACTGGAACACAAAGCTGAAGAAGAAGCTGATGAAAATGGGG
ATAGACCCGGTGACTCATAAACCGGTTTCTCAACTCCTTGACAGAACTTACAGAAACATTAGC
GGCCATGGAATGCATCCTTCAAAACAGAACCATCTAACAACCTTATACCTACACAATCC
AACTCAGCTTGGGAAATGATGAGAAACACAACAACAAACCATGAGAGTTATTACACCAAC
TCTCCAATGATGTTTACAATTCCTCTGAGTACCAAACTACTCCATTTTATTCTATAGC
CATCAAATCATCTGCTCAATGGAACCATCTTTCATGCTCTTCTCATCATCTTCTACT
AGTATCACTCAGCCAAACCAAGTACCTCAAACACCGGTTACTAAGTCTTACTGGAGCGAT
TTCCTTCTCTCGGACCCGGTTTCTCAAGTAGTGGGATCCTCAGCTACTAGCGACCTCACT
TTTACGAGAAGCAACATCAATTTCAACATCGAAGCCGAATACATCTCTCAAAACATCGAT
TCAAAGGCTCGGGAACATGTCTTCCGCGAGTTCTTCTGTTGACGAAATACTAGATAAA
GACCAAGAGATGTTGTACAGTTTCTCAACTCTTGAATGATTTGATTATTAG
>G1000 Amino Acid Sequence (domain in AA coordinates: 14-117)
MGRPPCCDKSNVKKGLWTEEDAKILAYVAIHGVGNWSLIPKKAGLNRCGKSCRLRWNY
LRPDLKHDSFSTQEEELIIECHRAIGSRWSSIARKLPGRDNDVKNHWNTKLKKLKMKG
IDPVTHKPVSQLAEFRNISGHGNASFKTEPSNNSILTSNSAWEMMRNTTNNHESYYTN
SPMMFTNSSEYQTTTFHFYSHPNHLLNGTSSSSSSSSSITQPNQVPQTPVTNFYWS
FLSDPVPQVVGSSATSDLTFTQNEHHFNIEAEYISQNDISKASGTCHSASSFVDEILDK
DQEMLSQFPQLLNDY*

>G1067 (436..1371)

TCTCAAGCTTCTCTCTCTTTTTCCTATAGCACATCAGAAATCGCTAAATACGACTCCT
ATGCAAGAAGAAGCTACTTCTTCTCTTGCCTAATTAATCTACCTAAGGGTTTCC
TCTTACCTTTCATGAGAGAGATCATTTAACAATAAGTCACCTTTTTATATCTTTTGCTTC
GTCTTTAATTTAGTCTCTGTTCTTGGTCTGTTTCTATATTTTGTGGCTTGCCTAACCGAT
CACACCTTAATGCTTTAGCTATTGTTTCTCAAAATCATGAGTTTGTGACTTCTCGATCTG
AGTTTCTTTTTCTCTCTTTACGCTCTTCTTCACTAGCTACCAATATATGAACGAGCAG
GATCAAGAATCGAGAAATTGATTTGAGCTGGCGAATAAGCAGTGGTGGGATAGGGAATTA
GTAGATGCGGCGCGGATGGAAGGCGGTACGAGCAAGGCGGTGGAGCTTCTAGATACTTC
CATAACCTCTTTAGACCGGAGATTCAACCAACAGCTTCAACCGCAGGGCGGGATCAAT
CTTATCGACAGCATCATCATCAGCACCAGCAACATCAACAACAACAACAACCGTCGGAT
GATTCAAGAGAATCTGACCATTTCAAACAAGATCATCATCAACAGGGTCCGACCGATCA
GACCCGAATACATCAAGCTCAGCACCAGGAAACGTCACGTTGACGTCACACGATCT
AAGAACAAGCCAAAGCCACCGATCATAGTAACTCGTGATAGCCCCAACCGCTTAGATCT
CACGTTCTTGAAGTATCTCTGAGCTGACATAGTTGAGAGTGTTCACGTACGCTAGG
AGGAGAGGAGAGGCGTCTCCGTTTTAGGAGGAAACGGCACCCTATCTAACGTCACTCTC
CGTCAGCCAGTCACTCTGGAATGGCGGTGGTGTGTCGGAGGAGGAGGAGTTGTGACT
TTACATGGAAGGTTTGAGATCTTTTCGCTAACGGGGACTGTTTTGCCACCTCCTGCACCG
CCTGGTGCCGCTGGTTTGTCTATATTTTAGCCGGAGGGCAAGGTCAGGTGGTGGGAGGA
AGCGTTGTGGCTCCCTTATTGCATCAGCTCCGTTATACTAATGGCGGCTTCGTTCTCA
AATGCGGTTTTCTGAGAGACTACCGATTGAGGAGGAGGAAGAAGAAGGTGGTGGTGGCGGA

GGAGGAGGAGGAGGAGGGCCACCGCAGATGCAACAAGCTCCATCAGCATCTCCGCCGTCT
GGAGTGACCGGTGAGGGACAGTTAGGAGGTAATGTGGGTGGTTATGGGTTTTCTGGTGAT
CCTCATTTGCTTGGATGGGGAGCTGGAACACCTTCAAGACCACCTTTTAAATTGAATTTT
AATGTCCGGAAATTTATGTGTTTTATCATCTTGAGGAGTCGTCTTTCCTTTGGGATATT
TGGTGTTTAATGTTTAGTTGATATGCATATTT

>G1067 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEGGYEQGGGASRYFHNLFPEIHHQQLQPQGGINLIDQHQQHQQHQQQQPSDDRES
DHSNKDHHQQGRPDSDPNTSSSAPGKRPRGRPPGSKNKAKPPIIVTRDSPNALRSHVLEV
SPGADIVESVSTYARRRGRGVSVLGGNGTVSNVTLRQPVTPGNGGGVSGGGVVTLHGRF
EILSLTGTVLPPPAPPGAGGLSIFLAGGQGVVGGSVVAPLIASAPVILMAASF SNAVFE
RLPIEEEEEEGGGGGGGGGGPPQMQQAPSASPPSGVTVGQQLGGNVGGYGFSGDPHLLG
WGAGTPSRPPF*

>G1075 (19..876)
TTTGTGTTTGGTGCTGGCATGGCTGGTCTCGATCTAGGCACAACCTTCTCGCTACGTCCAC
AACGTCGATGGTGGCGGCGGGACAGTTCACCACCGACAACCACCACGAAGATGACGGT
GGCGCTGGAGGAAACCACCATCATCACCATCATAATCACCATCAAGGTTTAGAT
TTAATAGCTTCTAATGATAACTCTGGACTAGGCGGCGGTGGAGGAGGAGGGAGCGGTGAC
CTCGTCATGCGTCGGCCACGTGGCCGTCCAGCTGGATCGAAGAACAACCGAAGCCGCCG
GTGATTGTACGCGCGAGAGCGCAAACTCTTAGGGCTCACATTCTGAAGTTGGAAGT
GGCTGCGACGTTTTCGAATGTATCTCCACTTACGCTCGTCGGAGACAGCGCGGATTTGC
GTTTATCCGGGACGGGAACCGTCACTAACGTCAGCATCCGTCAGCCTACGGCGGCCGGA
GCTGTTGTGACTCTGCGGGGTACTTTTGAGATTCTTTCCCTCTCCGGATCTTTCTTCCG
CCACCTGCTCCTCCAGGGGCGACTAGCTTGACGATATTCCTCGCTGGAGCTCAAGGACAG
GTCGTCGGAGGTAACGTAGTTGGTGAGTTAATGGCGGCGGGGCCGTAATGTCATGGCA
GCGTCTTTTACAAACGTGGCTTACGAAAGTTGCCTTTGGACGAGCATGAGGAGCACTTG
CAAAGTGGCGGCGGCGGAGGTGGAGGAATATGTACTCGGAAGCCACTGGCGGTGGCGGA
GGGTTGCCTTTCTTTAATTTGCGGATGAGTATGCCTCAGATTGGAGTTGAAAGTTGGCAG
GGGAATCAGCGCGGCGCGGTAGGGCTCCGTTTTAGCAATTTAAGAACTTTAATGTTTT
TTTCCACTTTTTGTTTTTCTCCGAATTTATGAAATTATGATTTAAGAAAAAACGAT
ATTGTTTCATGTATTGACCTCTTACTGCATGGTTTCTTCTATTGGGTTAATTGGCTAGCT
CATAAGAATTGTTTAAATTTGGTTATTGTTCATCAAATTTGCCACATATAAAGCTTCTAGC
AAAT

>G1075 Amino Acid Sequence (domain in AA coordinates: 78-85)
MAGLDLGTTSRYVHNVDGGGGQFTTDNHHEDDGGAGGNHHHHHHNHNHQQGLDLIASND
NSGLGGGGGGGSDLVMRPRGRPAGSKNKPPIVITRESANTLRAHILEVSGCDVFE
CISTYARRRQRCICVLSGTGTVNVSIRQPTAAGAVVTLRGTFEILSLSGSFLPPPAPPG
ATSLTIFLAGAQGVVGGNVVGMMAAGPVMMAASFNTVAYERLPLDEHEHLQSGGGG
GGGNMYSEATGGGGGLPFFNLPMSPQIGVESWQGNHAGAGRAPP*

>G1266 (62..718)
CAATCCACTAACGATCCCTAACCGAAAAACAGAGTAGTCAAGAAACAGAGTATTTTTCTA
CATGGATCCATTTTAAATTCAGTCCCATTCTCCGGCTTCTCACCAGGAATATTCTATCGG
ATCTTCTCCAGATTCTTCTCATCTCTTCTTCAACAATTACTCTCTTCCCTTCAACGA
GAACGACTCAGAGGAAATGTTTCTTACGGTCTAATCGAGCAGTCCACGCAACAAACCTA
TATTGACTCGGATAGTCAAGACCTTCCGATCAAATCCGTAAGCTCAAGAAAGTCAGAGAA
GTCTTACAGAGGCGTAAGACGACGGCCATGGGGGAAATTTCGCGGCGGAGATAAGAGATT
GACTAGAAACGGTATTAGGGTTTGGCTCGGGACGTTTCAAGCGCGGAAGAGCGGCTTT
AGCCTACGATCAAGCTGCTTCTCGATGAGAGGCTCTCGGCGATTCTCAATTTTTCGGC
GGAGAGAGTTCAAGAGTCGCTTTCGAGATTAAATATACCTACGAGGATGGTTGTCTCC
GGTTGTGGCGTTGAAGAGGAAACACTCGATGAGACGGAGAATGACCAATAAGAAGACGAA
AGATAGTGACTTTGATCACCCTCCGTGAAGTTAGATAATGTAGTTGTCTTTGAGGATTT
GGGAGAACAGTACCTTGAGGAGCTTTTGGGGTCTTCTGAAATAGTGGGACTTGTGAAA
GATTAGGATTTGTATTAGGGACCTTAAGTTTGAAGTGTTGATTAAATTTTAAACCCTAATA
TGTTTTTTGTTTGTCTTAAATATTGATTCTATTGAGAAACATCGAAAACAGTTTGTATGT
ACTTTTGTGATACTTGGCG

>G1266 Amino Acid Sequence (domain in AA coordinates: 79-147)
MDPFLIQSPFSGFSPEYSIGSSPDSFSSSSNNYSLPFNENDSEEMFLYGLIEQSTQQTY
IDSDSQDLPIKSVSSRKSEKSYRGVRRRPWGKFAAEIRDSTRNGIRVWLGTFFESAEEAAL

AYDQAAFSMRGSSAILNFSAEVQESLSEIKYTYEDGCSPPVALKRKHSMMRRMTNKKTK
DSDFDHRSVKLDNVVVFEDLGEQYLEELLGSSSENSGTW*

>G1311 (41..757)

AAGTATAATAACACAAAGAAACAGAGTAAAGAAAGAAAAATGGATTTTAAGAAGGAAGA
AACACTTCGTAGAGGGCCATGGCTCGAAGAAGAAGACGAACGGCTAGTGAAGGTCATTAG
TCTTTTGGGAGAACGTCGTTGGGATTCTTTAGCAATAGTTTCCGGTTTGAAGAGGAGTGG
TAAGAGTTGCAGGCTAAGGTGGATGAACCTATCTGAATCCGACTCTGAAGCGTGGACCGAT
GAGTCAAGAAGAAGAGAGAACAATTTTCAGCTCCATGCTCTATGGGGTAACAAGTGGTC
GAAGATTGCGAGAAGATTACCCGGTAGGACTGATAACGAGATAAAGAACTATTGGAGAAC
TCATTATAGAAAGAAACAGGAAGCTCAAACTATGGAAAGCTCTTTGAGTGGAGAGGAAA
TACAGGAGAAGAATTGTTGCACAAGTATAAGGAAACAGAGATCACTAGGACAAAGACGAC
GTCTCAAGAACATGGTTTTGTTGAAGTTGTGAGCATGGAAAGTGGTAAAGAAGCCAACGG
TGGTGTGGTGGGAAGAGAAAGCTTCGGTGTATGAAATCACCGTATGAAAATCGGATTTC
GGATTGGATATCAGAGATTTCTACTGACCAGAGTGAAGCAAATCTTTCAAGATCACAG
CAGCAATAGCTGCGAGTGAAGACAATATTAACTTGGTACTTGGTGGTTTTCAAGAGACTAG
GGACTTTGAGGAGTTTTCATGTTCTCTATGGTCATAATTCTAAAGTTGGTTTATTTACTT
TTTAAAAA

>G1311 Amino Acid Sequence (domain in AA coordinates: 11-112)

MDFKKEETLRRGPWLBEEDERLVKVISLLGERRWDSLAIIVSGLKRSKSCRLRWMNYLNP
TLKRGPMSQEEERIIFQLHALWGNKWSKIARRLPGRTDNEIKNYWRTHYRKQEAQNYGK
LFEWRGNTGEELLHKYKETEITRTKTTTSQEHGFVEVVSMSGKEANGGVGGRESFGVMKS
PYENRISDWISEISTDQSEANLSEDHSSNSCSENNINIGTWFPQETRDFFESCSSLWS*

>G1321 (72..803)

GTTCTTGATTGGTTTGGATCGGTATACTTAGTTGATTACGTAATTAAATAGATCGGCGT
GAAGAAGAAAAATGATCATGTGCAGCCGAGGCCATTGGAGACCAGCTGAAGACGAGAAGC
TCAAGGATCTTGTGCAACAATACGGTCCCTCACAATTGGAACGCCATTGCTCTCAAGCTTC
CTGGTCGCTCTGGTAAGAGTTGTAGATTGAGATGGTTTAATCAATTGGATCCAAGGATCA
ACCGAAACCCCTTTCACGGAAGAAGAAGAAAGACTTTTAGCGGCTCATCGGATCCATG
GGAACAGATGGTCCATCGCATCGCAAGGCTTTTCCCTGGAAGAACTGATAACGCCGTCAGA
ACCATGGCACGTCATCATGGCTCGTCGCACACGCCAAACCTCTAAGCCTCGTCTTCTTC
CCTCGACGACTTCGTCTTCTTTTAAATGGCGAGTGAACAAATCATGATGAGTTCTGGTG
GTTATAATCATAATTATAGTTCCGATGATCGGAAGAAAAATTTCCAGCAGACTTTATAA
ATTTCCCTTACAAATCTCTCATATCAATCATCTTCACTTCTTAAAGGAGTTTTTCCCG
GAAAGATCGCTTTAAGTCACAAAGCAAATCAGAGTAAGAAGCCTATGGAGTTCTACAATT
TTCTACAAGTAAACACAGATTCAAACAAGAGCGAGATTATAGATCAAGATTCAAGTCAAA
GCAACGCGAGTGAATCGGACACCAACATGAAAGTCATGTTCCATTCTTCGACTTTTAT
CCGTTGGAAACTCTGCCTCTAGGATTAGTTTTTTTGCAGTAACCTCTAAATTTCTAGAT
TAATATTTAGTCCGTATACGTACGAGATTATCTAGGTGCTTAGCATGTATGCTTGATGT
GTATAATCACTAAGTAGTGAGCTATTACCTGCGAAAATTGTAAGAAAAATACATAATGTT
GATGTATCACACATTCTCAATGTCTGTAAATTTCCATCGAGTTGTTAACTATCAAAGTT
ATCCGTTTGAAAAA

>G1321 Amino Acid Sequence (domain in AA coordinates: 4-106)

MIMCSRGHWRPAEDEKLKDLVEQYQPHNWNALKLPGRSKSCRLRWFNQLDPRINRNP
FTEEEERLLAAHRIHGNRWSIIARLFPGRTDNAVKNHWHVIMARRTRQTSKPRLLPSTT
SSSSLMASEQIMSSGGYNHNYSSDRKKIFPADFINFPYKFSHINHLHFLKEFFPGKIA
LSHKANQSKKPMFYNFLQVNTDSNKSEIIDQDSGQSKRSDSTKHESHVPFFDFLSVGN
SAS*

>G1326 (32..784)

CGACGGTACGGTGGAGATAGAGATAGCATCCATGGAGATGTCTAGAGGAAGCAACAGTTT
TGACAATAAGAAGCCTAGTTGCCAAAGAGGTCACTGGAGACCTGTTGAAGATGACAATCT
CCGGCAACTCGTTGAACAATACGGTCCCAAGAACTGGAATTTTATGCTCAACATCTCTA
TGGAAGATCAGGGAAGCTGTAGATTAAGATGGTACAACCAACTTGATCCAACATCAC
CAAGAAACCCCTTACCAGAGGAGGAAGAAGAGAGACTGCTTAAAGCTCATCGGATCCAAGG
GAATCGTTGGGCCCTCCATAGCCGACTGTTCCCCGGGAGGACCGACAACGCTGTCAAAA
CCATTTTCATGTATCATGGCTAGACGCAACGGGAAAACCTTCTTCCACAGCTACTTC
TACGTTCAACCAAACTTGGCATACTGTTTGAAGCCCTAGTTCTAGTCTTACAAGGCTAAA
TAGATCCCATTTCCGGCTATGGAGGTATCGAAAGGATAAGAGTTGCGGTCTCTGGCCTTA

CTCTTTTGTTCACCACCTACGAATGGTCAATTTGGATCTTCATCTGTCTCTAACGTACA
CCACGAAATTTATCTTGAGAGGAGAAAGTCGAAAGAGTTGGTGGATCCTCAGAATTACAC
ATTTTCATGCAGCCACACCAGATCATAAGATGACTTCAAATGAAGATGGACCATCCATGGG
AGATGATGGTGAGAAGAACGATGTTACTTTTCATTGATTTTCTTGGTGTGGATTAGCTTC
TTAGGTTATAACATCACAAGTCAAAGCTTTTAAGGGTTTCTATCATTAGGGTTAGGCATC
ATTTTCAGCCTTTTGCTTCCTTAAACTCTCATATGGATCT

>G1326 Amino Acid Sequence (domain in AA coordinates: 18-121)
MEMSRGNSFDNKKPSCQRGHWRPVEDDNLRLQVLEQYGPKNWNFIAQHLYGRSGKSCRLR
WYNQLDPNITKKPFTEEEERLLKAHRIQGNRWASIALRFPGRITDNAVKNHFHVMARRK
RENFSSTATSTFNQWHTVLSPPSSSLTRLNRSFGLWRYRKDKSCGLWPYSFVSPPTNGQ
FGSSSVSNVHIEIYLERRKSKELVDPQNYTFHAATPDHKMTSNEDGSPMGDDGEKNDVTF
IDFLGVGLAS*

>G1367 (128..1567)

TCCTTCCACAAAACCTTTTAAATTTTATCTGAAAAATTAACAACCCGAAACAAAAAAA
AAAATAAAAAATCAAAAAATCTCATCACCTTCCTTGCTCTGTATTTTCTCTCTCACTAA
ATCCTCCATGGATCCTTCTCTCTGCAACCAATGATCCTCATCATCCTCCTCCTCCTCA
GTTACATCTTTCCCTCCTTTCACCAACACCAACCCCTTCGCCCTCTCCAAACCACCCCTT
CTTCACCGGACCCACCGCCGTCGCGCCGCCAAACAACATCCATCTCTATCAAGCAGCTCC
TCCGCAGCAGCCACAAACATCTCCAGTTCCTCCTCATCCATCTATTTCCACCCCTCCTTA
CTCTGACATGATTTGCACGGCGATTGCAGCGTTAAACGAACCAGATGGGTCAAGCAAGCA
AGCTATTTTCGAGGTACATAGAGAGAATTTACACTGGGATTCCTACTGCTCATGGAGCTTT
GTTGACACACCATCTCAAGACTTTGAAGACCAGTGGGATTCCTGTCTATGGTTAAGAAATC
TTACAAGCTTGCTTCTACTCCTCCTCCTCCTCCTCCTACTAGTGTAGCTCCTAGTCTTGA
ACCTCCAGATCTGATTTTATAGTCAACGAGAACCAACCTTTACCTGATCCGGTTTGGC
TTCTTCTACTCCTCAGACTATTAAACGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT
TGTTGTTTCAACCTCAACCTCTGACTAATGGAAAACTCACCTGGGAACAGAGTGAATTACC
TGCTCTCTCGACCAGAGGAGATACAGATACAGCCGCCACAGTTACCGTTACAGCCACAGCA
GCCGGTTAAGAGACCGCCGGGTGCTCCTAGAAAAGATGGAACCTTCGCCGACGGTGAAGCC
AGCTGCTTCTGTTTCCGGTGGTGTGGAGACTGTGAAACGAAGAGGTAGACCTCCGAGTGG
AAGAGCTGCTGGGAGGGAGAGAAAGCCTATAGTAGTCTCAGCTCCAGCTTCAGTGTTCCT
GTATGTTGCTAATGGTGGTGTAGACGCCGAGGGAGACCAAGAGAGTTGACGCTGGTGG
TGCTTCTCTGTGTTGCTCCACCACCACCACCACTAACGTAGAGAGTGGAGGAGAGGA
GGTTGCAGTCAAGAAACGAGGAAGAGGACGGCCTCCTAAGATTGGAGGTGTTATCAGGAA
GCCTATGAAGCCGATGAGAAGCTTTGCTCGTACTGGAAAACCCGTAGGAAGACCCAGAAA
GAATGCCGTGTCTAGTGGGAGCTTCTGGACGACAAGATGGTGAATGAGGAACTGAAGAA
GAAGTTTGAGTTGTTTCAAGCGAGAGCTAAGGATATTGTAATTGTGTTGAAATCCGAGAT
AGGAGGAAGTGGAAATCAAGCAGTGGTTCAAGCCATACAGGACCTGGAAGGGATAGCAGA
GACAACAAACGAGCCAAAGCACATGGAAGAAGTGCAGCTGCCAGACGAGGAACACCTTGA
AACCAGAACGAGAAGCAGAGGGTCAAGGACAGACAGAAGCAGAGGCAATGCAAGAAGCTCT
GTTCTAAAGATAAAGCCTTGACATAAAAAGCTAGCAAGTGGTGGGTTTACTTGTGTGTG
TTACATGAAATTTTAAATCTTATAAGGGTGTGTCAGGAGAAAAACAAAAAGAACATGT
GATGAAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT
CTGTAAAGTGAATTAGGATGTTACCATGTTTCTGCTTCCATCTCTCTCCATCGTCCAT
ATCTGTGTAGGCAGCTTTGTTCTTTGTTCCCTCGTGTGTTTTTTTAGACTGTTGTGTCTCT
TATTCTATTTTGTCTCCTTAGGCTTTTTTAGGAGTTGTTGTTGATGTTTATCAAAAACGCT
TATGTAATTTTATGACCCTTCTACTTTTTATGATGGTTTCTT

>G1367 Amino Acid Sequence (domain in AA coordinates: 179-201, 262-285, 298-319,
335-357)

MDPSLSATNDPHHPPPPQFTSFPPFTNTNPFASPNHPFFTGPTAVAPPNNIHLHYQAAPPQ
QPQTSVPVPHPSISHPPYSDMICTAIAALNEPDGSSKQAIISRYIERIYTGIPHTAHGALLT
HHLKTLKTSGLVVMVKSKYKLASTPPPPPTSVAPSLPPRSDFIENENQPLPDPVLASS
TPQTIKRGRGRPPKAKPDVVQPQLTNGKLTWEQSELVSRPEEIQIQPPQLPLQPQPV
KRPPGRPRKDGTSPTVKPAASVSGGVETVKRRGRPPSGRAAGRERKPIVVSAPASVFPYV
ANGGVRRRGRPKRVDAGGASSVAPPPTTNVESGGEEVAVKKRGRGRPPKIGGVIRKPM
KPMRSFARTGKPVGRPRKNAVSVGASGRQDGYGELKKKFELEFQARAKDIVIVLKSEIGG
SGNQAVVQAIQDLEGIABETTNEPKHMEEVQLPDEEHLETEPEAEGQGQTEAEAMQEALF*
>G1386 (89..673)

AATTTTATTTCTTCTCTCAAATCTTCCCACCAAAAATTAACCTCTTTCGTTTCACACTAAG
 TCCCTTTTAAAAAGAAAATATCCCAATTAATGGAACGTGACGACTGCCGAGATTTTCAGGA
 CTCGCCGGCGCAGACGACGAGAGAGAAGAGTGAAATATAAAACCAAAGAAGAAAAGAGCCAA
 AGATGATGATGATGAGAAAAGTTGTTTTCGAAGCATCCAAATTTTCGAGGTGTCAGAATGAG
 ACAATGGGGAAAATGGGTGTCCGAAATCAGAGAGCCAAAAAAGAAATCAAGAATCTGGCT
 CGGTACTTTCTCCACGGCGGAGATGGCGGCGGTGCTCACGACGTGGCAGCTTTAGCCAT
 CAAAGGCGGTTCTGCACATCTCAACTTCCCAGGAGCTCGCTTATCACCTCCCTAGACCAGC
 TAGTGCCGACCTAAAGACATCCAAGCTGCCGCCGCCGAGCTGCAGCCGCTGTGGCCAT
 TGACATGGATGTAGAGACGTCTTCGCCGTGCCATCTCCCACAGTTACGGAAACGTATC
 TCCGCTATGATAGCACTCTCCGACGACGCTTCTCCGATCTTCTGATCTTCTGCTCAA
 CGTGAAACATAACATCGATGGCTTCTGGGACTCTTTTCCCTATGAAGAACCCTTCTCTC
 TCAAAGTTACTAGAACTCAAACTATGTCTGTTTTGTATGTATTTTGTCTATGTGACCA
 TTTTTTGACGTCGAAAATCACCCGATAATCCAAATTGTATGATTTATTAATGGTTGATG
 ATTTCTTTTGTGTGGAACAATGTGTATGATACGTAATCAAAAGTTCAAAAAAAAATAAA
 AAAAA

>G1386 Amino Acid Sequence (domain in AA coordinates: TBD)
 MERDDCRRFQDSPAQTERRVKYKPKKKRAKDDDEKVVSXHPNFRGVRMRQWGWVSEI
 REPKKKSRIWLGTFSTAEMAARAHDVAALAIKGSAGHLNFPDELAYHLPRPASADPKDIQA
 AAAAAAVIDMDVETSSPSPPTVTETSSPAMIALSDDAFSDLPDLLLNHNIDGFW
 DSFPYEEFPLSQSY*

>G1421 (292..1155)
 GAAATTCATCCCTAAATAAGAAAAAGCATCTCCTTCTTTAGTGTCTCTCTTCACCAAA
 CTCTTGATTCCATAAGCATATATTAAGAAAGCTCTCTGCTTTCTTCAACTTTCCCGGGAA
 AATCTTCTTGTACAAAGCATCAATCTCTTGTGTTTACCAATTTCTCTCTTTATTCCTTT
 TTTGCCCTTTACTTTTCTTAACCTTTGGTCTTTATATATAAACACACGACACAAAGAAGAA
 CACACATAAGTTAAACTATTACAACAGTTTTTAAAGAGAGAGATTTAAAAATGGAGACA
 GAGAAGAAAGTTTCTCTCCCAAGATCTTACGAATCTCTGTTACTGATCCTTACGCAACA
 GATTCTGTAAGCGACGAGAGAAGAAAGAGTTGATTTGATGCATTATCTACAAAACGACGT
 CGTGTTAAGAAGTACGTGAAGGAAGTGGTGCTTGATTCCGTGGTTTCTGATAAAGAGAAG
 CCGATGAAGAAGAAGAGAAGAAGCGCGTTGTTACTGTTCCAGTGGTTGTTACGACGGCG
 ACCGAGGAAGTTTCTGTTGAGTGAGGCAAGACCGTGGGGAAAATGGCGCGCGGAGATTAGA
 GATCCGAGTAGACGTGTTAGGGTTTGGTTAGGTACTTTTGACACGGCGGAGGAAGCTGCC
 ATTGTTTACGATAACGCAGCTATTACGCTACGTGGTCTTAACGCAGAGCTTAACCTCCCT
 CCTCTCCGGTGACGGAGAATGTTGAAGAAGCTTCGACGGAGGTGAAGAGAGTTTCGGAT
 TTTATCATTTGGCGGTGGAGAATGTCTTCGTTCCGCGGTTTCTGTTCTCGAATCTCCGTT
 TCCGGCGAGTCTACTGCGGTTAAAGAGGAGTTTGTGCGGTGATCGACGGCGGAGATTGTG
 GTTAAAAAGGAGCCGTCTTTTAACGGTTCAGATTTCTCGGCGCCGTGTTCTCGGACGAC
 GACGTTTTTGGTTTCTCGACGTGATGAGTGAAAGTTTCGGCGGCGATTTATTTGGAGAT
 AATCTTTTTCGGGATATGAGTTTGGATCCGGGTTTGGATTCCGGTCTGGGTCTGGATTC
 TCCAGCTGGCACGTTGAGGACATTTTCAAGATATTGGGGATTTATTCGGGTCCGATCCT
 GTCTTAACGTGTTTAAGAAATAACTGGCCGTTTAACGGCGTTTGTGAAAGTTTGTGTTACCG
 GCGACGGCGAGGATTAAAAAAAACGGCGATTTATTTTGAATGAAGATTGTTAAATA
 >G1421 Amino Acid Sequence (domain in AA coordinates: 74-151)
 METEKKVSLPRILRISVTDPYATDSSSDEEEVDFDALSTKRRRVKKYVKEVVLDSVSD
 KEKPMKKKKRKRVTVPVVVTATRKFRGRVQRPGWKWABIRDPSSRRVRVWLGTFTAE
 EAAIVYDAAIQLRGPNALNFPPTTENVEEASTEVKGVSDFIIGGGECLRSPVSVLE
 SPFGSESTAVKEEFVGVSTAEIVVKKEPSFNGSDFSAPLFSDDDVFGFSTSMSESFGGDL
 FGDNLFADMSFGSGFGSGSGFSSWHVEDHFQDIGDLFGSDPVLTV*

>G1453 (39..917)
 CGTCGACGCGAAATAAATCCTAGAAAATAACTATCAATATGATGAAGGTTGATCAAGATT
 ATTCTGTGTAGTATACCGCTGGATTTAGGTTTCATCCGACAGATGAAGAACTTGTCTGGAT
 ATTATCTCAAGAAGAAAATCGCTCCAGAGGATTGATCTCGACGTTATCAGAGAAATTG
 ATCTTTACAAGATCGAACCATGGGATCTACAAGAGAGATGTAGGATAGGGTACGAGGAGC
 AAACGGAGTGGTATTTCTTCAGCCATAGAGACAAGAAGTATCCGACTGGGACTAGGACAA
 ACCGAGCCACCGTGGCCGTTTCTGGAAAGCAACGGCCGGGACAAGGCGGTTTACCTCA
 ACTCAAACCTTATCGGTATGAGAAAAACGCTTGTCTTTTACCGAGGTCGAGCGCCTAATG
 GCCAAAAGTCCGATTGGATCATTCACGAATACTACAGCCTCGAGTCACACCAGAACTCTC

CTCCACAGGAAGAAGGATGGGTAGTGTGTAGAGCATTTAAGAAACGAACGACCATCCCAA
 CAAAAAGGAGGCAACTTTGGGATCCGAAC TGCTTATTCTACGACGACGCACTCTCTTGG
 AACCTCTCGACAAGCGAGCCAGACATAATCCTGATTTTACCGCCACACCGTTCAAGCAAG
 AACTACTCTCCGAGGCCAGTCACGTCCAGGATGGAGATTTCGGATCTATGTACCTTCAAT
 GCATCGATGATCAATTCTCCAGCTTCCTCAGCTCGAGAGCCCTCTCTCCGTCGG
 AAATAACTCCCCATAGTACTACTTTTCTGAGAACAGTAGCCGAAAGATGACATGAGCT
 CCGAGAAGAGGATCACTGACTGGAGATATCTAGATAAGTTCGTGGCGTCTCAATTTTGA
 TGAGTGGAGAAGACTAAAAAAGGCTTTCCTATGCATGCATGCACTAGAAACGTCGTGCGA
 TTTTGGATTTACATGCGGCCGCT

>G1453 Amino Acid Sequence (conserved domain in AA coordinates:13-160)
 MMKVDQDYSCSIPPGRFRFHPTDEELVGYYLKKKIASQRIDLVDVIREIDLYKIEPWLQER
 CRIGYEEQTEWYFFSHRDKKYPTGTRTNRATVAGFWKATGRDKAVYLNLSKLGMRKTLVF
 YRGRAPNGQKSDWIIHEYYSLESHQNSPPQEEGWVVCRAFKKRTTIPTKRRQLWDPNCLF
 YDDATLLEPLDKRARHNPDTATPFKQELLSEASHVQDGFSGMYLQCIDDDQFSQLPQL
 ESPSLPSEITPHSTTFSENSSRKDDMSSEKRITDWRYLDFKVASQFLMSGED*

>G1560 (120..1340)
 ATCCTTTCAATTTCCACTCTCTCTAATATAATTACATTTTCCCACTATTGCTGATTCA
 TTTTGTGTAATTATTTCAAACCCACATAAAAAATCTTTGTTTAAATTTAAACCA
 TGGATCCTTCATTTAGGTTTCATTAAAGAGGAGTTTCTGCTGGATTCACTGATTCTCCAT
 CACCACCATCTTCTTCTTCATACCTTTATTATCTTCCATGGCTGAAGCAGCCATAAATG
 ATCCAACAACATTGAGCTATCCACAACCATAGAAAGTCTCCATGAATCAGGGCCACCTC
 CATTTTGGACAAAGACATATGACTTGGTGGAAAGATTCAAGAACCAATCATGTGCTGTCTT
 GGAGCAAATCCAATAACAGCTTCATTGTCTGGGATCCACAGGCCCTTTCTGTAACCTCTCC
 TTCCAGATTTCTCAAGCACAATAACTTCTCCAGTTTGTCCGCCAGCTCAACACATATG
 GTTTCAGAAAGGTGAATCCGGATCGGTGGGAGTTTGCAAACGAAGGGTTTCTTAGAGGGC
 AAAAGCATCTCCTCAAGAACATAAGGAGAAGAAAAACAAGTAATAATAGTAATCAAATGC
 AACAACTCAAAGTTCTGAACAACAATCTCTAGACAATTTTGCATAGAAGTGGGTAGGT
 ACCGTCTAGATGGAGAGATGGACAGCCTAAGGCGAGACAAGCAAGTGTGATGATGGAGC
 TAGTGAGACTAAGACAGCAACAACAAAGCACCAAAATGTATCTCACATTGATTGAAGAGA
 AGCTCAAGAAGACCGAGTCAAAACAAAACAATGATGAGCTTCTTGCCTGCGCAATGC
 AGAATCCAGATTTTATTTCAGCAGCTAGTAGAGCAGAAGGAAAAGAGGAAAGAGATCGAAG
 AGGCGATCAGCAAGAAGAGACAAAGACCGATCGATCAAGGAAAAGAAATGTGGAAGATT
 ATGGTGATGAAAGTGGTTATGGGAATGATGTTGCAGCCTCATCTCAGCATTGATTGGTA
 TGAGTCAGGAATATACATATGGAACATGTCTGAATTCGAGATGTGCGAGTTGGACAAAC
 TTGCTATGCACATTCAAGGACTTGGAGATAATTCCAGTGCTAGGGAAGAAGTCTTGAATG
 TGGAAAAGGAAATGATGAGGAAGAAGTAGAAGATCAACAACAAGGGTACCATAAGGAGA
 ACAATGAGATTTATGGTGAAGGTTTTTGGGAAGATTGTTAAATGAAGGTCAAATTTTG
 ATTTTGAAGGAGATCAAGAAAATGTTGATGTGTTAATTGAGCAACTTGGTTATTTGGGTT
 CTAGTTACACACTAATTAAGAAGAAATGAAATGATGACTACTTTAAGCATTTGAATCA
 ACTTGTTTCTTATTAGTAATTTGGCTTTGTTTCAATCAAGTGAGTCGTGGACTAAGTGC
 >G1560 Amino Acid Sequence (domain in AA coordinates: 62-151)
 MDPSFRFIKEEFPAGFSDSPSPSSSSSYLYSSSMAEAAINDPTTLSYPPQPLEGLHESGPP
 PFLTCKTYDLVEDSRNTHVSVWSKSNNSFIWDPQAFSVTLPRFFKHNNFSSFVRQLNTY
 GFRKVNPRDRWEFANEGFLRGQKHLKKNIRRRKTSNNSNQMQPQSSEQQSLDNFCIEVGR
 YGLDGEMLSLRRDKQVLMELVRLRQQQSTKMYLTLIEEKLKKTESKQKQMSFLARAM
 QNPDFIQQLVQKEKRKEIEEAISKRRQRPIDQGRNVEDYGDESGYNDVAASSALIG
 MSQEYTYGNMSEFEMSELDKLAHIIQGLGDNSSAREEVLNVEKGNDEBEVEDQQQGYHKE
 NNEIYEGGFWE DLLNEGQNFDFEGDQENVDLIQQLGYLGSSSHTN*

>G1594 (1..984)
 ATGGATGGAATGTACAATTTCCATTCCGCCGGTGATTATTCAGATAAGTCGGTTCTGATG
 ATGTCACCGGAGAGTCTCATGTTTCTTCCGATTACCAAGCTTTGCTATGTTCTCCGCC
 GGTGAAAATCGTGTCTCTGATGTTTTCGGATCCGACGAGCTACTCTCAGTAGCCGTCTCC
 GCTTGTGTCGTCCGAGCGCTTCGATCGCTCCGAGATCCGAAGAAATGATGATAACGTT
 TCTCTAACTGTCATCAAAGCTAAAATCGCTTGTATCCTTCGTATCCTCGCTTACTTCAA
 GCTTACATCGATTGCCAAAAGGTCCGAGCACCACCGAGATAGCGTGTCTTACTAGAGGAG
 ATTCAACGGGAGAGTGATGTTTATAAGCAAGAGGTTGTTCTTCTTCTTGGAGCT
 GATCTGAGCTTGATGAATTTATGGAAACGTACTGCGATATATTAGTGAAATACAAATCG

GATCTAGCAAGACCGTTTGACGAGGCAACGTGTTTCTTGAACAAGATTGAGATGCAGCTA
 CGGAACCTATGTACTGGTGTGAGTCTGCCAGGGGAGTTTCTGAGGATGGTGTAATATCA
 TCTGACGAGGAAGTGAAGGTGATCATGAGGTAGCAGAGGATGGGAGACAAAGATGT
 GAAGACCGGGACCTCAAAGATAGGTTGCTACGCAAATTTGGAAGCCGTATTAGTACTTTA
 AAGCTTGAGTTCTCAAAGAAGAAGAAGAAAGGAAAGTTACCAAGAGAAGCAAGACAAGCT
 CTCTTGATTGGTGAATCTCCATTATAAGTGGCCTTACCCTACTGAAGGAGATAAGATA
 GCATTAGCTGATGCAACGGGGTTAGACCAAAAAACAAATCAACAATTGGTTTATAAACCAA
 AGGAAACGTCATTGGAAGCCATCAGAGAATATGCCTTTTCGCTATGATGGATGATTCTAGT
 GGATCATTCTTTACCGAGGAATGA

>G1594 Amino Acid Sequence (conserved domain in AA coordinates: 343-308)

MDGMYNFHSAGDYSKSVLMMSPELSMFPSPDYQALLCSSAGENRVSDVFGSDELLSVAVS
 ALSSEAASIAPEIRNDDNVS LTVIKAKIACHPSYPRLLQAYIDCQKVGAPPEIACLEEE
 IQRESDVYKQEVVPSSCFGADPELDEFMETYCDILVKYKSDLARPFDEATCFLNKIEMQL
 RNLCTGVESARGVSEDGVISSDEELSGGDHEVAEDGRQRCEDRDLKDRLLRKFPSRISTL
 KLEFSKKKKKGKLPREARQALLDWNWNLHYKWYPTEGDKIALADATGLDQKQINNWFINO
 RKRHWKPSENMPFAMDDSSGSFFTEE*

>G1750 (94..1101)

CCCTTTTCCTCTCTTCTCCAAATCTCTGAAAATTTTACCAGAATCTCTGTTCTTTTTT
 TCACCAGAATCTCTCTGTTTAAATAATAGGTGATGATGATGGATGAGTTTATGGATCTT
 AGACCAGTGAAGTACACAGAGCACAAAGACTGTTATCAGAAAGTACACTAAAAAGTCGTCT
 ATGGAGAGGAAGACCAAGTGTTCGTGACTCGGCCAGGTTGGTTTCGGGTCTCAATGACGGAT
 CGTGACGCCACTGATTCATCAAGCGACGAGGAAGAGTTTCTGTTCCCTCGAAGACGTGTC
 AAGAGATTGATTAAACGAGATCAGAGTCGAGCCTAGCAGCTCTTCCACCGGCGACGTCTCT
 GCTTCTCCGACGAAGGACCGGAAAAGAATCAACGTTGATTCTACGGTTCAAAGCCCTCT
 GTTTCGGGCCAAAACCGAGAAGAAGTACCGCGGCGTGAGACAGCGACCATGGGGAAAATGG
 GCGGCGGAGATTCTGTATCCTGAGCAACGCCGAGAAATCTGGCTCGGTACTTTTGCAACG
 GCGGAGGAAGCTGCCATCGTCTACGACAACGACGCAATCAAACCTTCGTGGCCCTGATGCT
 CTTACCAACTTCACCGTACAACCGAACCAGAACCAGAACCAGTACAAGAACAAGAACCAG
 GAGAGCAACATGTCTGGTTTCGATATCAGAATCAATGGACGATTCTCAACATCTATCATCT
 CCGACATCGGTTCTCAACTACCAACATATGTCTCGGAGGAACCAATCGATAGTCTTATC
 AAACCGGTTAAACAAGAGTTTCTTGAACCAGAACAGAGCCAATAGCTGGCATCTTGGA
 GAAGGTAATACTAATACTAATGATGATTCAATTTCCATTGGACATTACATTTCTCGACAAC
 TATTTCAATGAATCATTACCAGACATCTCCATCTTCGATCAACCTATGTCTCCTATTCAA
 CCAACAGAGAATGATTTCTTCAACGACCTTATGTTATTCGATAGCAACGAGAAGAATAC
 TACTCTCCGAGATCAAAGAGATTGGTTTCATCGTTCAACGATCTTGATGATTCTTTGATA
 TCCGATCTCTTACTTGTGTGATATTTTGGCCATTAACCAAACACCGGTTTGGTTGC

>G1750 Amino Acid Sequence (domain in AA coordinates: 107-173)

MMMDEFMDLRPVKYTEHKIVIRKYTKKSSMERKTSVRDSARLVRVSMTRDARDSSDEE
 EFLFPRRRVKRLINEIRVEPSSSSTGDVSASPTKDRKRINVDSTVQKPSVSGQNKYRG
 VRQRPWGKWAAEIRDPEQRRRIWLGTFATAEEAAIVYDNAAIKLRGPDALTNFTVQPEPE
 PVQEQQEPESNMVSISSEMDDSQHLSSPTSVLNYQTYVSEEPIDSLIKPVKQEFLEPE
 QEPISWHLGEGNTNTNDDSFPLDITFLDNYFNESLPDISIFDQPMSPIQPTENDFFNDLM
 LFDNSNAEYYSSSEIKEIGSSFNDLDDSLISDLLLV*

>G1947 (70..918)

ACAACTATTCTCTCTCTCTCTTTTTTTTATTA AAAAAGCTCAAATTTATATAGGTTTTTT
 GTTCACAAAATGGATTATAACCTTCCAATTCCATTAGAGGGTCTCAAAGAAACGCCACCA
 ACGGCTTTCTTGACGAAAACATACAACATAGTGGAGGATTCAAGCACAAACAACATAGTT
 TCATGGAGCAGAGAACAACAACAGCTTCATTGTTTGGGAACCAGAGACTTTTGGCCCTAATT
 TGCTTCCCTAGATGCTTTAAGCACATAATTTCTCCAGCTTTGTTAGACAGCTCAATACT
 TATGGGTTTAAAGAAGATTGATACAGAGAGATGGGAATTTGCAAATGAGCATTTTCTGAAG
 GGAGAGAGGCATCTTCTTAAGAACATCAAGAGAAGAAAGACATCATCTCAAACGCAAACG
 CAGTCGCTAGAAGGAGAGATCCATGAGCTGCGAAGAGACAGAATGGCTTTAGAAGTAGAA
 CTGGTTAGACTGCGACGAAAACAAGAAAGCGTGAAGACATATCTGCATTTGATGGAAGAG
 AAAGTGAAGTCAAGAAAGTAAAGCAAGAAATGATGATGAATTTCTTGCTAAAGAAGATT
 AAGAAACCGAGTTTTTTACAGAGCTTAAGGAAACGTAATCTGCAAGGAATCAAGAATCGA
 GAGCAAAAGCAAGAGGTGATCTCAAGCCATGGTGTGAGGATAATGGAAAGTTTGTATAA
 GCTGAGCCAGAAGAGATGTTGATGACATCGATGATCAATGTGGAGGTGTGTTTGATTAT

GGTGATGAGCTTCACATAGCTTCAATGGAGCATCAAGGACAAGGGGAGGATGAAATTGAA
 ATGGATAGTGAAGGAATTTGGAAGGGTTTCGTGTTGAGTGAGGAGGAGATGTGTGATTGA
 GTGGAACATTTTATATAATAAACTAATGTATTATGAGAGGTTTTTTTTTTGTTTTTTGCT
 TTTTTTTTCCGAGTTTGTCAAGCATGTATACAATTGGGGCCAACTAAAAGCCCAA
 CAAAATATTGGCCTTGGCATTGTGTTAACAAATTGACTAATTCGGCCACACCTTCC

>G1947 Amino Acid Sequence (domain in AA coordinates: 37-120)
 MDYNLPIPLEGLKETPPTAFLTKTYNIVEDSSTNNIVSWSRDNNFIVWEPETFALICLP
 RCFKHNNFSSFVRQLNTYGFKKIDTERWEFANEHFLKGERHLLKNIKRRKTSSQTQTQSL
 EGEIHELRRDRMALEVELVRLRRKQESVKTYLHLMEEKLKVTEVKQEMMMNLLKKIKKP
 SFLQSLRKRLNQGIKNREQKQEVISSHGVEDNGKFVKAEPPEYGGDDIDDQCGGVFDYGD
 LHIASMEHQQGEDEIEMDSEGIWKFVLSSEEMCDLVEHFI*

>G2011 (309..1547)

AATGTCGGTTGTACAATTATTTGTCACTAAAGTTTCCAAATTTCTTCTAAACTGATGAAT
 CAATGGAACATGATGACGAAAAAGATAAATCCACGGTGGCGGGAACGACCCACCCATT
 CCACCGCTCTCTATTTCCAGATTTTTTCAATTATCTGACTACAGTTTGTCTCGTTACT
 TCCTTCCCTAAACCTTTATAAACCATTAACCTCTCATCTTCTTCTTAAACCCCTA
 ATTATCACACACACCCCAATTTCTCACTCTCTCTCACTAAAACCCGTAAATTTTCTAC
 TATATCAAATGAGCCCAAAAAAGATGCTGTTTCTAAACCAACTCCAATTTCACTACCCG
 TTTGAGAGCATCCGATATACCCGGGTCTCTTACGTCGACACTGACATGGGTTTCTCTG
 GGTCAACACTTCCCATGCCACTAGACATCTTACAAGGGAATCCAATTTCCACCTTTTTTAT
 CCAAGACTTTTGATTGGTTGATGACCCGACTCTTGACCCGGTCATCTTGGGGACTGA
 CCGGAGCTAGCTTCGTAGTTTGGGATCCTTAGAGTTTGCCAGAATCATACTTCCAAGGA
 ATTTCAAACACAACAATTTCTCCAGCTTCGTCAGACAGCTTAACACTTATGGATTTTGGAA
 AGATTGATACTGACAAGTGGGAATTCGCTAACGAGGCTTTCTTAGAGGCAAGAAGCATC
 TTCTGAAGAACATTATCGTCGTCGATCACCACAATCCAACCAAACTTGCTGCAGTAGCA
 CTAGCCAAAGCAAGGGTCACCTACTGAGGTTGGAGGAGAGATTGAGAAGCTGAGGAAAG
 AGCGGCGTGCATTGATGGAGGAAATGGTTGAGCTTCAGCAGCAAAGCAGAGGCACAGCTC
 GACATGTGGACACTGTAAACCAGAGGCTGAAAGCTGCAGAGCAACGTGAGAAGCAATTGC
 TCTCTTTCTTGGCTAAGTTGTTTCAAGACCGGGTTTCTTGAACGCCTGAAGAATTCA
 AAGGAAAAGAAAAGAGGAGCTCTTGGATTGGAAAAGGCGAGAAAAGATTATCAAGC
 ACCACCAGCAGCTCAAGATTCTCCAACAGGAGGGGAGGTGGTGAAGTATGAAGCTGATG
 ATTGGGAGAGATTGCTAATGTATGACGAAGAGACTGAGAACACCAAGGGTTTAGGAGGGA
 TGACTTCAAGCGATCCAAAAGGCAAGAATTGATGTATCCATCAGAAGAAGAGATGAGCA
 AACCAGATTACTTGATGTCTTCCCATCTCTGAAGGACTTATTAACAAGAAGAGACGA
 CATGGAGCATGGGTTTCGATACTACAATACCGAGTTTCAGCAACACCGATGCATGGGGAA
 ACACAATGGACTATAATGATGTCTCAGAGTTTGGTTTGTCTGCAGAAACAACAAGTGATG
 GTTTCCTGATGTCTGTGCGGAACAATTTGCTGCAAGGAATCAGAGAGACTGGATTCAACT
 GGCCAACCTGGTGATGATGATGATAATACGCCAATGAATGATCCTTAGGATCTTTTCATAT
 ATAGTTTAGACCAAAAACCCGTTTCTTATCGGGTGAACATTAATTCAATTATTCATTTTG
 AATGCACTCTTTATACATATATATAATATTGATGAGTTTGATTGTTCCAAAAA

>G2011 Amino Acid Sequence (domain in AA coordinates: 56-147)
 MSPKKDAVSKPTPIVPSVSRSDIPGSLYVDITDMGFSGLPLMPLDILQGNPIPPFLSKT
 FDLVDDPTLDPVISWGLTGASFVWDPLEFARIILPRNFKHNNFSSFVRQLNTYGFRRKID
 TDKWEFANEAFRLRGKKHLLKNIHRRRSPQSNQTCCSSTSQSGSPTEVGGIEKLRKERR
 ALMEEMVELQQQSRGTARHVDTVNQRLKAAEQRQKQLLSFLAKLFQNRGFLERLKNFKGK
 EKGGALGLEKARKKFIKHQQPDSPGTGGEVVKYEADDWERLLMYDEETENTKGLGGMTS
 SDPKGKNLMYPSEEMSKPDYLSFSPSEGLIKQEETWSMGFDTTIPSPSNTDAWGNTM
 DYNDVSEFGFAAETTS DGLPDVCWEQFAAGITETGFNWP TGDDDDNTPMNDP*

>G2094 (1..450)

ATGCTAGATCCCAACCGAGAAAGTAATCGATTGAGAAATCAATGGAAGCAAACCTCACATCA
 GTAGATGCGATCGAAGAACACAGCAGTAGCAGTAATGAAGCTATCAGCAACGAGAAG
 AAGAGTTGTGCCATTTGTGGTACCAGCAAAACCCCTCTTTGGCGAGGCGGTCCTGCCGGT
 CCCAAGTCGCTTTGTAACGCATGCGGGATCAGAAACAGAAAGAAAGAAGAACTGATC
 TCAATAGATCAGAAGATAAGAAGAAGAAGATCATAACAGAAACCCGAAGTTTGGTGAC
 TCGTTGAAGCAGCGATTAAATGGAATTGGGGAGAGAAGTGATGATGACGCGATCAACGGCT
 GAGAATCAACGGCGGAATAAGCTTGGCGAAGAAGAGCAAGCCCGCTGTACTCATGGCT
 CTCTCTTATGCTTCTTCCGTTTATGCTTAA

>G2094 Amino Acid Sequence (domain in AA coordinates:43-68)
MLDPTEKVIDSESMESKLTSDVAIEEHSSSSSNEAISNEKKSCAICGTSKTPLRGGPAG
PKSLCNACGIRNRKKRRTLISNRSEDKKKKSHNRNPKFGDSLKQRLMELGREVMMQRSTA
ENQRRNKLGEEEQAAVLLMALSYASSVYA*

>G2113 (90..590)
ATAACAAACTCATCAAACTTCCTCAGCGTTTCTTTTCTTACATAAAACAATTTTCTTAC
ATAACAAATCTTGTGTTTGTGTTGTCATGGCACCACAGTTAAACCGCGGCCGTCA
AAACCAACGAAGGTAACGGATCCGTTACAGAGGAGTGAGGAAGAGACCATGGGGACGTT
ACGCAGCCGAGATCAGAGATCCTTTCAAGAAGTCACGTGTCTGGCTCGGTACTTTCGACA
CTCCTGAAGAAGCCGCTCGTGCCTACGACAAACGTGCTATTGAGTTTCGTGGAGCTAAAG
CCAAAACCAACTTCCCTTGTTCACATCAACGCCCACTGCTTGAGTTTGACACAGAGCC
TGAGCCAGAGCAGCACCGTGGAATCATCGTTTCTTAATCTCAACCTCGGATCTGACTCTG
TTAGTTTCGAGATTCCCTTTTCTAAGATTACGGTTAAGGCTGGGATGATGGTGTTCGATG
AAAGGAGTGAATCGGATCTTCGTGGTGATGGATGTCGTTAGATATGAAGGACGAC
GTGTGGTTTGGACTTGGATCTTAATTTCCCTCCTCCACCTGAGAACTGATTAAAGATTTA
ATTATGATTATTAGATATAATTAATGTTTCTGAATTGAG

>G2113 Amino Acid Sequence (domain in AA coordinates: TBD)
MAPTVKTA AVKTNEGNVRYRGRVGRYPWGRYAAEIRDPFKSRVWLGTFTDTPEEAARAYD
KRAIEFRGAKAKTNFPYCYNINAHCLSLTQSLSQSSTVSESPNPNLNGSDSVSSRFPFKI
QVKAGMMVFDERSSESSSVMDVVRVYEGRRVLDLNLNFPFPPEN*

>G2115 (41..733)
AATCACTCTACAAAGCCTGTACGTACACAACAACATTACCATGGTGAAACAAGAACGCAA
GATCCAAACCAGCAGCACAAAAAGGAAATGCCTTTGTTCATCATCACCATCTTCTTCTTC
TTCTTCATCTTCTTCTCGTCTTCGTCTTCGTGTAAGAACAAGAACAAGAAGAGTAAGAT
TAAGAAGTACAAAGGAGTGAGGATGAGAAGTTGGGGATCATGGGTCTCTGAGATTAGGGC
ACCAATCAAAAGACAAGGATTGGTTAGGTTCTTACTCAACAGCTGAAGCAGCTGCTAG
AGCTTACGATGTTGCACTCTTATGTCTCAAAGGCCCTCAAGCCAATCTCAACTTCCCTAC
TTCTTCTTCTCTCATCATCTTCTTGATAATCTCTTAGATGAAAATACCCTTTTGTCCCC
CAATCCATCCAAAGAGTAGCTGCTCAAGCTGCCAACTCATTAAACCATTTTGCCCCCTAC
TTCATCAGCCGTCTCGTCAACGCTCCGATCATGATCATCACCATGATGATGGGATGCAATC
TTTGATGGGATCTTTTGTGGACAATCATGTGCTTTGATGGATCAACATCTTCATGGTA
TGATGATCATAATGGGATGTTCTTGTGTTGATAATGGAGCTCCATTCAATTACTCTCCTCA
ACTAAACTCGACGACGATGCTCGATGAATACTTCTACGAAGATGCTGACATTCGGCTTTG
GAGTTTCAATTAATCCGACGGTCCATAATACATACTTTAATTAGT

>G2115 Amino Acid Sequence (conserved domain in AA coordinates:46-115)
MVKQERKIQTSSSTKEMPLSSSSSSSSSSSSSSCKNKNKSKIKYKGVMRMSWGS
WVSEIRAPNQKTRIWLSYSTAEAAARAYDVALLCLKGPQANLNFPTSSSSHLLDNLDD
ENTLLSPKSIQRVAAQAANSFNHFAPTSSAVSSPSDHDHDDGMQSLMGSFVDNHVSLM
DSTSSWYDDHNGMFLFDNGAPFNYSPLNSTTMLDEYFYEDADIPLWSFN*

>G2130 (41..988)
CCTCTCTTCATTTTTTAACTCCCTCTCTCTCTCTCTCTATGGAGAGACGAACGAGACG
AGTGAAGTTCACAGAGAATCGTACGGTCACAAACGTAGCAGCTACACCATCTAACGGGTC
TCCGAGACTGGTCCGTATCACTGTTACTGATCCTTTTCGCTACTGACTCGTCTAGCGACGA
CGACGACAACAACAGTCACGGTGGTTCCAAGAGTGAAACGATACGTGAAGGAGATTAG
ATTCTGCCAAGGTGAATCTTCTTCTCCACCGCGGAGGAAAGGTAAGCACAAGGAGGA
GGAAAGCGTAGTGGTTGAAGATGACGTGTCGACGTGCGTGAAGCCTAAAAAGTACAGAGG
CGTGAGACAGAGACCTTGGGGAAATTCGCGGCGGAGATTAGAGATCCGTCGAGCCGTAC
TCGGATTTGGCTTGGGACTTTTGTACGGCGGAGGAAGCTGCTATAGCGTACGATAGAGC
CGCGATTCTCTCAAAGGACCTAAAGCGCTCACGAATTTCTTAACCTCCGCCGACGCCAAC
CCGGTTATCGATCTCCTCAACCGGTTTCCGCTGCGATTACGGTAGAGATTCTCGGCAGAG
CCTTCATTACCCGACCTCTGTTCTAAGATTCAACGTCAACGAGGAAACAGAGCATGAGAT
TGAAGCGATCGAGCTATCTCCGAGAGAAAGTCGACGGTTATAAAGAAGAAGAAGAATC
GTCGGCGGGTTTGGTGTTCGGGATCCGTATCTGTTACCGGATTATCTCTCGCCGCGCA
ATGTTTTTGGGATACCGAAATGCCCCCTGACCTTTTGTCTTCGATGAAGAAACCAAAAT
CCAATCAACGTTGTTACCAACACAGAGGTTTCGAAACAAGGAGAAACGAAACTGAAGA
TTTCGAGTTTGGTTGATTGATGATTTTCGAGTCTTCTCCATGGGATGTGGATCATTTCTT
CGACCATCATCACTCTTTCGATTAAAAATCTCTCTTTTGGGGAAATTTTGTG

>G2130 Amino Acid Sequence (domain in AA coordinates 93-160)
MERRTRRVKFTENRTVTNVAATPSNGSPRLVRITVTDPFATDSSSDDDNNVTVVPRVK
RYVKEIRFCQGESSSSTAARKGKHKEESVVVEDDVSTSVKPKKYRGVRQRPWGKFAAEI
RDPSSRTRIWLGTFTVTAEEAAIAYDRAAIHLKGPALTNFLTPPTPVIDLQTVSACDY
GRDSRQSLHSPTSVLRFNVNEETEHEIEAIELSPERKSTVIKEEEESSAGLVFPDPYLLP
DLSLAGECFWDTEIAPDLLFLDEETKIQSTLLPNTTEVSKQGENETEDFEFGLIDDFESSP
WDVDHFFDHHHHSFD*

>G2147 (162..1262)
CTGTGATTGTCAAGAGTTTGAACACACAAAGAAGAAAGAAGAACTCAACATTTCAAGCAA
GAAGAAAGAGAGAAGAGAGAAGGTCCAATAATAGAGAGAACAAAAAAGAGAGCTTAA
TTGTCAAGTTTATTCTCTGCAACGTGCGGCCTAAGTAACACATGTCTGAATTATGGAGTTA
AAGAGCTCACATGGGAAATGGGCAACTAACCGTTCATGGTCTAGGCGACGAAGTAGAAC
CAACCACCTCGAATAACCTATTGGACTCAAAGTCTCAACGGTGTGTGAGACTTTGGAGT
CTGTGGTTCATCAAGCGGCTCTACAGCAGCCAAGCAAGTTTCAGCTGCAGAGTCCGAATG
GTCCAAACCCACAATTATGAGAGCAAGGATGGATCTTGTTCAGAAAACCGGTTATCCTC
AAGAAATGGACCGATGGTTCGCTGTTCAGAGGAGAGCCATAGAGTTGGCCACAGCGTCA
CTGCAAGTGCAGTGGTACCAATATGTCTTGGGCGTCTTTTGAATCCGGTTCGGAGCTTGA
AGACAGCTAGAACCAGAGACAGAGACTATTCCGCTCTGGATCGGAACTCAAGATACTG
AAGGAGATGAACAAGAGACAAGAGGAGAAGCAGGTAGATCTAATGGACGACGGGACGAG
CAGCAGCGATTACAACGAGTCCGAAAGGAGACGGCGTGATAGGATAAACAGAGGATGA
GAACACTTCAGAAGCTGCTTCTTCTACTGCAAGTAAGGCGGATAAAGTCTCAATCTTGATG
ATGTTATCGAACACTTGAAACAGCTACAAGCAAGTACAGTTCATGAGCCTAAGAGCCA
ACTTGCCACAACAAATGATGATTCCGCAACTACCTCCACCACAGTCAGTTCTCAGCATCC
AACACCAACAACAACAACAACAACAGCAGCAGCAGCAGCAACAACAGCAGCAACAGTTTC
AGATGTCTGTTGCTTGCAACAATGGCAAGAATGGGAATGGGAGTGGTGGAAATGGTTATG
GAGGTTTAGTTCCTCTCTCTCTCTCTCCACCAATGATGGTCCCTCTCTATGGGTAACAGAG
ACTGCACCAACGGTCTTTCAGCCACATTATCTGATCCATACAGCGCTTTTTTCGCACAGA
CAATGAATATGGATCTCTCAATAAATGGCAGCAGCTATCTATAGACAACAGTCTGATC
AAACAACAAGGTAATATCGGCATGCCTTCAAGTCTTTCGAATCATGAGAAAAGAGATT
AGTCTAGCGACCTAGTATTATTGATCCATATATATAGTTCTTGAAAGATTGTTGTATCAT
GATTGTAAAAACTGTTTGTAGTATGGAAAAAGACTTGCAGATAAAA

>G2147 Amino Acid Sequence (domain in AA coordinates:160-234)
MSNYGVKELTWENGQLTVHGLGDEVEPTTSNNPIWTQSLNGCETLESVVHQAALQQPSKF
QLQSPNGPNHNYESKDGSCSRKRGYPQEMDRWFVAVQEESHVRVGHVSTASASGTNMSWASF
ESGRSLKTARTGDRDYFRSGSETQDTEGDEQETRGEAGRSNGRRGRAAAIHNESERRRRD
RINQRMRTLQKLLPTASKADKVSILDDVIEHLKQLQAQVQFMSLRANLPQQMMIPQLPPP
QSVLSIQHQQQQQQQQQQQQQQQQQFQMSLLATMARMGMGGGGNGYGLVPPPPPPPMV
PPMGNRDCNTNGSSATLSDPYSAFFAQTMMNDLYNKMAAAIYRQSDQTTKVNIGMPSSSS
NHEKRD*

>G2156 (384..1292)
TTTTTTTTCCCTTTCCTCGTTCAAAAAAGTACTTGCAGAGTCACTCACTCTCAGTCTCA
GCACATGAATTAATTTGAAGCTTCCCTAGAATTCTTTACATCAATTAATACGACACCGT
CTCGGGTGAAGAATCTCTCTCTCTTGCCTAAAGCGAGTTAGGGTTTAAACACACAAAGC
ATACCTTTTAGATTGTGTCTCTTAGCTCTGTTTTGTGCGCTTGTGTAACCGATCAACT
CAAGCTATTGGTCTCTCACCTCTGAAATTTGACTTCTCCAATGGATCTCAAAGTTTCTC
TTATATGAATCTATCTTACCCTCACAATATCTTTATATATATGAGCCACAAGAACAAG
AAGAGTCAGTAGATGCGGCTGCCATGGACGGTGGTTACGATCAATCCGGAGGAGCTTCTA
GATACTTTCAACACTCTTCAGGCTTGAGCTTCATCACCAGCTTCAACCTCAGCCTCAAC
TTACCTTTTGCCTCAGCCTCAGCCTCAACCTCAGCCTCAGCAGCAGAATTCAGATGATG
AATCTGACTCCAACAAGGATCCGGGTTCCGACCCAGTTACCTCTGGTTCAACCGGGAAC
GTCCACGTGGACGTCTCCGGGATCCAAGAACAAGCCGAAGCCACCGGTGATAGTACTA
GAGATAGCCCCAACGTGCTTAGATCTCATGTTCTTGAAGTCTCATCTGGAGCCGACATAG
TCGAGAGCGTTACCACTTACGCTCGCAGGAGAGGAAGAGGAGTCTCCATTCTCAGTGGTA
ACGGCACGGTGGCTAACGTGATCTCCGGCAGCCGGCAACAGACGCGGCTCATGGGGCAA
ATGGTGAACCGGAGGTGTTGTGGCTCTACATGGAAGGTTTGTAGATACTTTCCCTCAGAG
GTACGGTGTGTCGGCCCTTGCAGCCGAGGATCCGGTGGTCTTTCTATCTTTCTTCCG
CGCTTCAAGGTGAGGTGATGGAGGAAACGTGGTGGCTCCGCTTGTGGCTTCGGGTCCAG

AACGGTGAAGGCCGATGGAATCTGCTCGCTAAATCTTCTGGGCTAAAGAGAGCAGGAAAA
 AGTTGTAGATTGAGATGGTTGAATTACCTTAAACCCGACATAAAGCGTGGGAATCTCACT
 CCTCAAGAACAACCTTTTAACTCTTGAGCTCCATTCTAAATGGGGTAATAGGTGGTCAAAA
 ATTTTCGAAGTATTTACCAGGAAGAACAGACAACGATATCAAAAACCTACTGGAGAACTAGA
 GTCCAGAAACAAGCACGCCAGCTCAACATAGATTCCAATAGCCACAAGTTCATAGAAGTT
 GTTCGTAGCTTTTGGTTTCCAAGACTGATCAACGAGATTAAAGACAACCTCATAACCAAC
 AATATTAAGCTAATGCTCCTGATTTACTTGGACCAATTTTACGAGACAGCAAAGATTTG
 GGTTCACAAACATGGATTGTTCACCTTCCATGTGAGAAGATCTCAAGAAAACCTTCACAA
 TTCATGGATTTTCTGATCTTGAAACCACATGTCTTGGAGGATCACGAGGGGGTAGT
 AGTCAATGTGTGAGTGAGTTTATAGCTCCTTCCCTTGCCTAGAGGAGGAGTACATGGTG
 GCCGTTATGGGCAGTTTACAGCATTTTACGATTGTCATGATTGTACGTGGCTGATTCCAAG
 TACGAGGATGATGTGACACAAGATCTAATGTGGAACATGGATGACATTTGGCAGTTTAAAC
 GAGTATGCACACTTTAATTAGGTTATATTATTTATGTACTTCTTACAACCTGGAGGGG
 TTTATCGGTCTTTTATTAAATTTGATTGTTTTGGATTCTTAAAAATGTGTTCTTATTA
 TAGTTTTTAAAGAAAAAATGTTTAAAGCGCAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
 >G2893 Amino Acid Sequence (conserved domain in AA coordinates: 19-120)
 MSNITKKKCNNGNEEGAEQRKGPWLTLEEDTLLTNYISHNGEGRWNLLAKSSGLKRAGKSCR
 LRWLNLYLKPDIKRNLTLPQEQLLILELHLSKWNRSKISKYLPGRTDNDIKNYWRTRVQK
 QARQLNIDSNSHGFIEVVRSFWRFLINEIKDNSYTNNIKANAPDLLGPILRDSKDLGFN
 NMDCSTSMSEDLKKTQFMDFDLETTMSLEGRSGSSQCVSEVYSSFPCLLEEYMVAVM
 GSSDISALHDCHVADSKYEDDVTQDLMWNMDDIWQFNEYAHFN*
 >G340 (97..834)
 ATGAAATCTCTGTAGTTTTTTTTTGTTCCTTTCTTAAATTTTCAAAGAAAGACATTTATT
 AAACCAAAATAACTCTTTAGATCATTGCAAGGAAAAATGTTGAAAAGTGCAAGTCCAATG
 GCATTCTACGATATCGGAGAGCAGCAATACTCTACTTTCGGGTACATTTTAAGCAAACCT
 GGGAAACGACAGGAGCTTACGAGATTGACCTTCGATCCCAAACATCGACGATGCGATCTAC
 GGCTCAGATGAGTTCCGTATGTACGCTTACAAAATCAAACGGTGTCTCTGACTCGTAGC
 CACGACTGGACGGAGTGTCCCTACGCTCACCCTGGCGAGAAAGCCACACGCCGTGATCCT
 CGCCGTTACACTTACTGTGAGTGCATGCCCGGCTTTCGAAATGGCGCATGCCACCGT
 GCGGACTCATGCGAATTTCGCACATGGCGTATTCGAGTACTGGCTCCACCCGGCGCGTTAC
 CGAACACGCGCATGTAACGCCGGGAACCTGTGTGAGAGGAAAGTGTGTTTCTTTGCCAC
 GCGCCGGAGCAGCTAAGGCAGTCTGAAGGAAAGCACAGGTGCAGGTACGCATATAGGCCG
 GTGAGGGCTAGAGGTGGTGGAAACGGCGATGGAGTGACGATGAGAATGGACGACGAGGGT
 TACGACACGTACCGTCTCCGGTGAGAAGCGGGAAAGATGATTTAGATAGTAACGAGGAG
 AAGGTGTTGTTGAAGTGTGGAGTCGGATGAGCATTGTGGATGATCATTATGAGCCGTCC
 GATTTGGATTTGGATTGTGACACTTTGATTGGATCTCAGAGTTGGTTCGATTAAATTTGG
 GAAATCAAAGCAGAGAACAAGAAACCCGATAAATAAAGTGGATTTTGTAAATCCAC
 AAGATCAAGATTCAAGATGAGAGATCTTGTGATGTATATGGTAAATTTAATTGTAATGAT
 TTATTGCAATGTCGCAAAAGAAGTTACTTCTCTTTCATGTAAACAGATTCTTGATCTTC
 TATAAGTCTTTGTATTAA
 >G340 Amino Acid Sequence (domain in AA coordinates: 37-154)
 MLKSASPMFYDIGEQYSTFGYILSKPGNAGAYEIDPSIPNIDDAIYGSDEFMYAYKI
 KRCPRTRSHDWTECPYAHREKATRRDPRRYTYCAVACPFRNGACHRGDSCEFAHGVFE
 YWLHPARYRTRACNAGNLCQRKVCFFAHAPEQLRQSEGKHRCRYAYRPVRARGGGNGDGV
 TMRMDDEGYDTSRSPVRSGKDDLSNEEKVLLKCSRMSIVDDHYEPSDLDDLSDLHFDWI
 SELVD*
 >G39 (75..638)
 GTTTCACAGTCCCGTACTTGTGCATAAACTGTAAAACTACTCTGAAAATTTTGCT
 TCTGTTAGGATATAATGCCACCCTCTCCTCTAAATCTCCTTTTATTAGCTCTTCACTCA
 AAGGAGCTCATGAAGATCGCAAATTTAAATGCTATAGGGGTGTCCGAAAGAGGTCTTGGG
 GCAAATGGGTGTCTGAAATCAGAGTTCCAAAGACTGGACGACGAATATGGCTAGGTTTAT
 ACGATGCTCCAGAGAAGGCAGCTAGAGCCTATGATGCTGCTTTGTTCTGTATTAGGGGTG
 AGAAGGGAGTTTACAATTTTCCACTGATAAAAAGCCGAGCTTCCAGAAGGTTCTGTCC
 GGCTCTGTCCAAAGCTCGACATACAGACAATAGCAACAACTATGCTTCATCAGTTGTGC
 ATGTACCTTCCCATGCCACCACACTCCCGGCAACAACCCAGGTTCCCTCTGAAGTTCCTG
 CTTCCTCTGATGTTTCTGCTTCTACTGAGATTACAGAGATGGTTCGATGAATATTATCTCC
 CAACCGATGCAACTGCAGAATCAATATTCTCAGTTGAAGACTTACAACCTGGACAGTTTCC

TCATGATGGACATTTGATTGGATAAAACAATCTAATCTGATGTGTAACGTCACCTTGCAGTGA
CATTTAATATGGTTTANCTATCAGTTACCTGTCTGCTTCTTGTAAAGGTATACTTGGATC
CTTGCTTTTGAACCTTGTTTTATTTAGCATGCAAA

>G39 Amino Acid Sequence (domain in AA coordinates: 24-90)

MPPSPKSPFISSSLKGAHEDRKFKCYRGVRKRSWGWSEIRVPKTGRRIWLGSYDAPE
KAARAYDAALFCIRGEKGVYNFPTDKKPQLPEGSVRPLSKLDIQTATNYASSVVHVP
SH
ATTLPATTPQVPSEVPASSDVSASTEITEMVDEYYLPTDATAESIFSVEDLQLDSFLMMDI
DWINNLI*

>G439 (128..967)

TATAAATCTTCGTTTCTACTTTTTTTTTCTTCCATAATATAGTCAATTCGTTTTCTTAATT
AGGGCTTCTTCTCTTTGTTTCTCCAATCTTATTAGTTTATTTATTTTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATTCATGAAGGTAACCTCATCTTCTTCTACTTCGAATTCATCAAATCCAAAACC
ATTAACCTCTAATTTTCAATCCCTAATAATGACCAAGTCTTACCGGTATCTAACC
TCCGATTGGGCTAAACCAAGCTCACTCCAACACAAATCCTCCAAATTCAGACAGAGTTACA
TCTCCGGCAAAACCAATCTCGTCGTCGCGTGGTAGTCATCTTCTCACCCTAAACCAAC
CTCAATGAAGAAAATCGACGTAGCAACTAAACCGGTTAACTATACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTCCGGCTACCTAAAAACCGAACCCGGTTATG
GCTCGGTACGTTTCAAGACGGCTCAAGAAGCTGCATTAGCTTACGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTGTTTCAAGGACACTATAA
ACAGATATTGTCCTCCGTCTATCAACGCAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCCTCAGATCGAGAAACAGAAACAAAACAGAGGAGGTCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCGGAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGCACAAGTATCCTTCTTTGGAGATTGATTGGGACGCTATAGAGAACT
CTTCTGAATCCATTTTATCTTTTGGATTGATTGCTCTAAATTTGTAATTTTATTTTTC
AGAGCTTTGTAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGAACCTAACTCT
GTTTTCTTTTGTAAAGTATTGTTTATAATGGGCCGTTGAATGGGCCCTTATTGATTAAACA
GCCCAAGTTTTTAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)

MAMALNMNAYVDEFMEALEPFMKVTSSTSSNSNPKPLTPNFIPNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPVKLYRGVRQRQ
WGKWVABIRLPKNRTRLWLGTFFETAQEAALAYDQAAHKIRGDNARLNFDPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEEVLSGFSKPEKEPEFGIYCGYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G470 (1..2580)

ATGGCGAGTTCCGGAGTTTCAATGAAAGGTAATCGTGGAGGAGATAACTTCTCCTCTCT
GGTTTTAGTGACCCTAAGGAGACTAGAAATGTCTCCGTGCGCCGGCGAGGGGCAAAAAGT
AATTCTACCCGATCCGCTGCGGCTGAGCGTGCTTTGGACCCTGAGGCTGCTCTTTACAGA
GAGCTATGGCAGCCTTGTGCTGGTCCGCTTGTGACGGTTCCTAGACAAGACGACCGAGTC
TTCATTTTCTCAAGGACACATCGAGCAGGTGGAGGCTTCGACGAACCGCGGCAGAA
CAACAGATGCCCTCTATGATCTTCCGTCAAAGCTTCTCTGTCGAGTTATTAATGTAGAT
TTAAAGGCAGAGGCAGATACAGATGAAGTTTATGCGCAGATTACTCTTCTCCTGAGGCT
AATCAAGACGAGAATGCAATTGAGAAAGAAGCGCTTCTCTCCACCTCCGAGGTTCCAG
GTGCATTGCTTCTGCAAAACCTTGACTGCATCCGACACAAGTACACATGGTGGATTTTCT
GTTCTTAGGCGACATGCGGATGAATGTCTCCACCTCTGGATATGTCTCGACAGCCTCCC
ACTCAAGAGTTAGTTGCAAAAGGATTTGCATGCAATGAGTGGCGATTGAGACATATATTC
CGGGGTCAACCACGAGGCAATTTGCTACAGAGTGGGTGGAGTGTGTTTGTAGCTCCAAA
AGGCTAGTTGACAGGCGATGCGTTTATATTTCTAAGGGGCGAGAATGGAGAATTAAGAGTT
GGTGTAAAGCGTGCGATGCGACAACAAGGAACGTCGCGTCTTCTGTTATATCTAGCCAT
AGCATGCATCTTGGAGTACTGGCCACCCGATGGCATGCCATTTCAACAGGGACTATGTTT
ACAGTCTACTACAAACCCAGGACGAGCCCATCTGAGTTTATTGTTCCGTTGATCAGTAT
ATGGAGTCTGTTAAGAATAACTACTCTATTGGCATGAGATTCAAAATGAGATTTGAAGGC
GAAGAGGCTCCTGAGCAGAGGTTTACTGGCACAATCGTTGGGATTGAAGAGTCTGATCCT
ACTAGGTGGCCAAAATCAAAGTGGAGATCCCTCAAGGTGAGATGGGATGAGACTTCTAGT
ATTCCTCGACCTGATAGAGTATCTCCGTGGAAGTAGAGCCAGCTCTTGCTCCTCCTGCT
TTGAGTCTCTGTTCCAATGCCTAGGCCCTAAGAGGCCAGATCAAATATAGCACCTTCATCT

CCTGACTCTTCGATGCTTACCAGAGAAGGTACAACCTAAGGCAAACATGGACCCCTTTACCA
GCAAGCGGACTTTCAAGGGTCTTGCAAGGTCAAGAATACTCGACCTTGAGGACGAAACAT
ACTGAGAGTGTAGAGTGTGATGCTCCTGAGAATTCTGTTGTCTGGCAATCTTCAGCGGAT
GATGATAAGGTTGACGTGGTTTCGGGTTCTAGAAGATATGGATCTGAGAACTGGATGTCC
TCAGCCAGGCATGAACCTACTTACACAGATTGCTCTCCGGCTTTGGGACTAACATAGAT
CCATCCCATGGTCAGCGGATACCTTTTTATGACCATTATCATCATCACCTTCTATGCCTGCA
AAGAGAATCTTGAGTGATTGAGAAGGCAAGTTCGATTATCTTGCTAACCACTGGCAGATG
ATACACTCTGTCTCTCCCTGAAGTTACATGAATCTCCTAAGGTACCTGCAGCAACTGAT
GCGTCTCTCCAAGGGCGATGCAATGTTAAATACAGCGAATATCCTGTTCTTAATGGTCTA
TCGACTGAGAATGCTGGTGGTAACTGGCCATACGTCCACGTGCTTTGAATTATTATGAG
GAAGTGGTCAATGCTCAAGCGCAAGCTCAGGCTAGGGAGCAAGTAACAAAACAACCCCTC
ACGATACAAGAGGAGACAGCAAGTCAAGAGAAGGGAAGTGCAGGCTCTTTGGCATTCCT
CTGACCAACAACATGAATGGGACAGACTCAACCATGTCTCAGAGAAACAACCTTGAATGAT
GCTGCGGGGCTTACACAGATAGCATCACCAAGGTTTCAGGACCTTTCAGATCAGTCAAAA
GGGTCAAAATCAACAAACGATCATCGTGAACAGGGAAGACCATTCCAGACTAATAATCCT
CATCCGAAGGATGCTCAACGAAAACCAACTCAAGTAGGAGTTGCACAAAGGTTCACAAG
CAGGGAATTGCACTTGGCCGTTCACTGGATCTTTCAAAGTTCCAAAACATAGAGGAGTTA
GTCTGCTGAGCTGGACAGGCTGTTTGTAGTTCAATGGAGAGTTGATGGCTCCTAAGAAAGAT
TGGTTGATAGTTTACACAGATGAAGAGAATGATATGATGCTTGTGTTGGTGACGATCCTTGG
CAGGAGTTTGTGTTGCATGGTTTCGCAAAATCTTCATATACAGAAAGAGGAAGTGAGGAAG
ATGAACCCGGGGACTTTAAGCTGTAGGAGCGAGGAAGAAGCAGTTGTTGGGGAAGGATCA
GATGCAAGGACGCCAAGTCTGCATCAAACTCTTCATTGTCCAGCGCTGGGAACCTTTAA
>G470 Amino Acid Sequence (domain in AA coordinates: 61-393)
MASSEVSMKGNRGGDNFSSSGFSDPKETRNVS VAGEGQKSNSTRSAAERALDPEAALYR
ELWHACAGPLVTVPRQDDRVFYFPQGHIEQVEASTNQAAEQMPLYDLPSKLLCRVINVD
LKAEDTDEVYAQITLLPEANQDENAIEKEAPLPPPPRFQVHSFCKTLTASDSTHGGFS
VLRHHADECLPPLDMSRQPPTQELVAKDLHANEWRFRIFRGQPRRHLLQSGWSVVFSSK
RLVAGDAFIFLRGENGELRVGVRRAMRQQGNVPSSVISHSMLGLVLATAWHAISTGTMF
TVYKPRTPSPSEFIVPFDQYMESVKNNYSIGMRFKMRFEGEAEPEQRFTGTIVGIEESDP
TRWPKSKWRSRLKVRWDETSSIPRPDRVSPWKVEPALAPPALSPVPMRPRKPRSNIPSS
PDSSMLTREGTTKANMDPLPASGLSRVLQGGYEYSLRTKHTESVECDAPENSVVWQSSAD
DDKVDVVSGSRRYGSSENWMSARHEPTYTDLGSGFGTNIDPSHGQRIPFYDHSSSPSMPA
KRILSDSEGKFDYLANQWQMIHSGLSLKLHESPKVPAATDASLQGRCNVYSEYPVLNGL
STENAGGNWPIRPRALNYEEVVNAQAQAQAREQVTKQPFITQEETAKSREGNCRFLFGIP
LTNNMNGTDSTMSQRNNDAAAGLTQIASPKVQDLSQSKGSKSTNDHREQGRPFQTNPNP
HPKDAQTKTNSSRSCTKVHKQGIAGRSVDLSKFQNYEELVAELDRLFEFNGELMAPKKD
WLIVYTDEENDMMLVGDDPWQEFCCMVRKIFIYTKEEVRKMNPGTTLSCRSEEEAVVGEES
DAKDAKSASNPSSLSSAGNS*
>G652 (1..606)
atgagcggaggagacgtgaacatgagtggtggagacagacgcaaggggaacggtgaag
tggtttgatacacagaaggggtttggtttcatcacacctagcagcgggtggtgacgatctc
ttcggttcaccagtcctccatcagatctgaaggatttcgtagcctcgagctgaggaatct
ggtgagttcgacgttgaggttgacaactccggcgtcccaaggctattgaagtgtctgga
ccgacggtgctcccggttcagggtaacagcgggtggtggtggttcatctggtggacgcggt
ggttttgccggcggtggtggaagaggagggggacgtggtggaggaagctacggaggaggt
tatggtggaagaggaagcgggtggcgtggaggaggtggtggtgataattcttgctttaag
tgcggtgaaccaggtcacatggcgagagaatgctctcaaggtggtggaggatacagcggga
ggcgggggtggtggaaggtacgggtctggcggcgggcggaggaggaggtggtggtggctta
agctgctacagctgtggagagctctgggcactttgcaagggattgcactagcgggtggtgct
cgttga
>G652 Amino Acid Sequence (domain in AA coordinates: 28-49, 137-151, 182-196)
MSGGDDVNMSGGDRRKGTWKVFDTPQKGFITPSDGGDDL FVHQSSIRSEGFRLAAEES
VEFDVEVDNSGRPKAIEVSGPDGAPVQGNSSGGSSSGGRGGFGGGGGGRGGGGGGGGG
YGGRGSGGRGGGGGDNCFKCEGPHMARECSQGGGGYSGGGGGGGRYSGGGGGGGGGGGL
SCYSCGESGHFARDCTSGGAR*
>G671 (61..1119)
TTCACTTGAGAACACCCCTTTGAACTCGATCAAGAAAGCTAAGTTTGAAGAATCAAGA

ATGGTGCGGACACCGTGTGCAAAGCCGAACTAGGGTTAAAGAAAGGAGCTTGGACTCCC
GAGGAAGATCAGAAGCTTCTCTTACCTTAACCGCCACGGTGAAGGTGGATGGCGAACT
CTCCCCGAAAAGCTGGACTCAAGAGATGCGGCAAAAGCTGCAGACTGAGATGGGCCAAT
TATCTTAGACCTGACATCAAAAGAGGAGAGTTCACTGAAGACGAAGAACGTTCAATCATC
TCTCTTCACGCCCTTCACGGCAACAAATGGTCTGCTATAGCTCGTGGACTACCAGGAAGA
ACCGATAACGAGATCAAGAACTACTGGAACACTCATATCAAAAAACGTTTGATCAAGAAA
GGTATTGATCCAGTTACACACAAGGGCATAACCTCCGGTACCGACAAATCAGAAAACCTC
CCGGAGAAACAAAATGTTAATCTGACAACCTAGTGACCATGATCTTGATAATGACAAGGCG
AAGAAGAACAACAAGAATTTTGGATTATCATCGGCTAGTTTCTTGAACAAAGTAGCTAAT
AGGTTTCGGAAAGAGAATCAATCAGAGTGTTCGTCTGAGATTATCGGAAGTGGAGGCCCA
CTTGCTTCTACTAGTCACACTACTAATACTACAACCTACAAGTGTTCCTGACTCTGAA
TCAGTTAAGTCAACGAGTTCTTCCCTTCGCACCAACCTCGAATCTTCTGCCATGGGACC
GTTGCAACAACACCAAGTTTCATCGAAGTTTGACGTTGATGGTAACGTTAATCTGACGTGT
TCTTCGTCCACGTTCTCTGATTCTCCGTTAACAATCCTCTAATGTACTGCGATAATTTT
GTTGGTAATAACAACGTTGATGATGAGGATACATCGGGTTCTCCACATTTCTGAATGAT
GAAGATTTTATGATGTTGGAGGAGTCTTGTGTTGAAAAACACTGCGTTTATGAAAGAACTT
ACGAGGTTTCTTCACGAGGATGAAAACGACGTCGTTGATGTGACGCCGGTCTATGAACGT
CAAGACTTGTGTTGACGAAATGATAACTATTTTGGATGAGTGAAGTCAATATCGATGAA
TCCCACGTGACCATGTCAATATGATGTCTATGGATATGTTACCTTGATGATGTTGATGGT
AATAATAATAAATAATAATAGATGGTGATGATGACCATGCATGAATCATGAATGTAGTTCGTG
TTGTACATATGCTTGTGTTTTTGTGTTTTTTTTTTTGGTCTGAAGTGTGTTGTTTCGT
TGTAATGGATTATAAATGGTGATGTAATAATTATAATGTTAAAAA

AAAA

>G671 Amino Acid Sequence (domain in AA coordinates: 15-115)
MVRTPCCKAELGLKKGAWTPEEDQKLISYLNHRHGEWWRTLPEKAGLKRCGKSCRLRWAN
YLRPDIKRGEFTEDEERSIISLHHLHGNKWSAIARGLPGRDNEIKNYWNTHIKRLIKK
GIDPVTHKGITSGTDKSENLPKQNVNLTSDHDLNDNDKAKNNKNFGLSSASFLNKVAN
RFGKRINQSVLSEIIGSGGPLASTSHTTNTTTTSVSVDSSESVKSTSSSFAPTSNLLCHGT
VATTPVSSNFDVGNVNLTCSSSTFSDSSVNNPLMYCDNFVGNVNVDEEDTIGFSTFLND
EDFMMLEESCVENTAFMKELTRFLHEDENDVVDVTPVYERQDLFDEIDNYF*

>G779 (110..712)

GACATGCATGTAAGCATTCGGTTAATTAATCGAGTCAAAGATATATATCAGTAAATACAT
ATGTGTATATTTCTGGAAAAAGAAATATATATATTGAGAAATAAGAAAAGATGAAAATGGA
AAATGGTATGTATATAAAAGAAAGGAGTGTGCGACTCTTGTGTCTCGTCCAAAAGCAGATC
CAACCAACAGCCCCAAAAGAAGCATGATGGAGCCTCAGCCTCACCATCTCCTCATGGATTG
GAACAAAGCTAATGATCTTCTCACACAAGAACACGCAGCTTTTCTCAATGATCCTCACCA
TCTCATGTTAGATCCACCTCCCGAAACCTAATTTCACTTGGACGAAGACGAAGAGTACGA
TGAAGACATGGATGCGATGAAGGAGATGCAGTACATGATCGCCGTATGCAGCCCGTAGA
CATCGACCCTGCCACGGTCCCTAAGCCGAACCGCGTAACGTAAGGATAAGCGACGATCC
TCAGACGGTGGTTGCTCGTCCGGCTCGGGAAAGGATCAGCGAGAAGATCCGAATTTCTCAA
GAGGATCGTGCCTGGTGGTGAAGATGGACACAGCTTCCATGCTCGACGAAGCCATACG
TTACACCAAGTTCTTGAACCGGCAGGTGAGGATTCTTCAGCCTCACTCTCAGATTGGAGC
TCCTATGGCTAACCCCTCTTACCTTTGTTATTACCACAACCTCCCAACCCTGATGAACCTAC
ACAGAAGCTCGCTAGCTAGACATTTGGTGTCATCTCTCAACCTTT

>G779 Amino Acid Sequence (domain in AA coordinates: 126-182)
MKMENGMYKKKGVCDSCVSSKSRSNHSPKRSMMEPQPHLLMDWNKANDLLTQEHAAFLN
DPHLLMLDPPPETLIHLDEDEEYDEMDAMKEMQYMIAVMQPVDIDPATVPKPNRRNVRI
SDDPQTIVARRRRERISEKIRILKRIVPGAKMDTASMLDEAIRYTKFLKQVRILQPHS
QIGAPMANPSYLCYYHNSQP*

>G962 (148..1392)

CGTCTGACTCTCTACTCAACACCACTCAATTTTCATCTCTCTTTTCCCTTCCATTGTTAGT
ATAAAAACCAAGCAAACCTTAATCACTTTTCATCATCATATATCACCTTAATCCCATG
CATACACATATCTAGTCTTTTGGATATATGGCAATTGTATCCTCCACAACAAGCATCATT
CCCATGAGTAACCAAGTCAACAATAACGAAAAAGGTATAGAAGACAATGATCATAGAGGC
GGCCAAGAGAGTCATGTCCAAAATGAAGATGAAGCTGATGATCATGATCATGACATGGTC
ATGCCCGGATTTAGATTCCATCTACCGAAGAAGAACTCATAGAGTTTACCTTCGCGCA
AAAGTTGAAGGCAAACGCTTTAATGTAGAACTCATCACTTTCCTCGATCTTTATCGCTAT

GATCCTTGGGAACCTCCTGCTATGGCGGCGATAGGAGAGAAAGAGTGGTACTTCTATGTG
CCAAGAGATCGGAAATATAGAAATGGAGATAGACCGAACCGAGTAACGACTTCAGGATAT
TGGAAAGCCACCGGAGCTGATAGGATGATCAGATCGGAGACTTCTCGGCTATCGGATTA
AAGAAAACCCCTAGTTTTCTACTCTGGTAAAGCCCCCTAAAGGCACCTCGTACTAGTTGGATC
ATGAACGAGTATCGTCTTCCGACCATGAAACCGAGAAGTACCAAAAGGCTGAAATATCA
TTGTGCCGAGTGTAACAAAGGCCAGGAGTAGAAGATCATCCATCGGTACCACGTTCTCTC
TCCACAAGACATCATAACCATAACTCATCGACATCATCCGTTTAGCCTTAAGACAACAA
CAACACCATTTCCTCTCTAATCATTCCGACAACAACCTTAACAACAACAACAACATC
ACAATCTCGAGAAGCTCTCCACCGAATATTCCGGCGACGGCAGCACAACAACAGACC
ACAAACAGTAACCTGACGTTACCATTGCTCTAGCCAATCAAAACATATATCGTCCAATG
CCTTACGACACAAGCAACAACACATTGATAGTCTTACGAGAAATCATCAAGACGATGAT
GAACTGCCATTGTTGACGATCTTCAAAGACTAGTTAACTACCATAATCAGATGGAGGT
AACATCAATCACCATACTTTCAAATGCTCAACAGTTTCATCATACTCAACAACAAAAT
GCTAACGCAACGCATTACAATTGGTGGCTGCGGCGACTACAGCGACAACGCTAATGCCT
CAAACTCAAGCGCGCTTAGCTATGAACATGATTCTGCGAGGAACGATTCCAAACAATGCT
TTGTGGGATATGTGAATCCAATAGTACCAGATGGAAACAGAGATCACTATACATAATATT
CCTTTTAAGTAATTTAATTAGATCATGATTATTATCCATGACAATAATTAATGCTGCTTT
CGCG

>G962 Amino Acid Sequence (domain in AA coordinates: 53-175)
MAIVSSTTSIIPMSNQVNNNEKGIENDHRGGQESHVQNEDEADDHDDHDMVMPGFRFHPT
EEELIEFYLRKVEGKRFNVELITFLDLRYDPWELPAMAAIGKEWYFYVPRDRKYRNG
DRPNRVTTSGYWKATGADRMIRSETSRPIGLKTLVFPYSGKAPKGTRTSWIMNEYRLPHH
ETEKYQKAEISLCRVYKRPVEDHPSVPRSLSTRHHNHSSTSSRLALRQQQHSSSSNH
SDNNLNNNNINNNLEKLSTEYSGDGSTTTTTTNSNSDVTIALANQNIYRPMPTDTSNNTL
IVSTRNHQDDDETAIVDDLQRLVNYQISDGGNINHQYFQIAQQFHHTQQQNANANALQLV
AAATTATTLMPQTQAALAMNMI PAGTI PNNALWDMWNPIVPDGNRDHYTNIPFK*
>G977 (46..591)

CACCAAACCTCACCTGAAACCCATTTCCATTACCATTACACTAATGGCACGACCACAA
CAACGCTTTTCGAGGCGTTAGACAGAGGCATTGGGGCTCTTGGGTCTCCGAAATTCGTCAC
CCTCTCTTGAACAAAGAATCTGGCTAGGGACGTTTGAGACAGCGGAGGATGCAGCAAGG
GCCTACGACGAGGCGGCTAGGCTAATGTGTGGCCCGAGAGCTCGTACTAATTTCCCATAC
AACCTAATGCCATTCTACTTCTCTTCCAAGCTTCTATCAGCAACTCTTACCGCTAAA
CTCCACAAATGCTACATGGCTTCTCTTCAAATGACCAAGCAAACGCAACACAAACGCAA
ACGCAGACCGCAAGATCACAATCCGCGGACAGTGACGGTGTGACGGCTAACGAAAGTCAT
TTGAACAGAGGAGTAACGGAGACGACAGAGATCAAGTGGGAAGATGGAATGCGAATATG
CAACAGAATTTTAGGCCATTGGAGGAAGATCATATCGAGCAAATGATTGAGGAGCTGCTT
CACTACGGTTCCATTGAGCTTTGCTCTGTTTTACCAACTCAGACGCTGTGAGAAATGGCC
TTGTCTGTTTAGCGTATCTTTTCAATTTTATTTTTGTTTCCACAAAAACGGCGTCGTAA
GTGATGAGAGTAGTAGTGAGAGAAGGCTAATTTCAAGACATTTTGATCTGAATTGGCCTC
TTTTGAAACACTGATTCTAGTTTCTATAAGAGCAATCGATCATATGCTATGTTATGTATA
GTATTATAAAAAAATGTTATTTTCTGATTNAAAAAAAAAAAAAAAAAAAAAAAAA
>G977 Amino Acid Sequence (domain in AA coordinates: 5-72)

MARPQQRFRGVRQRHWGSWVSEIRHPLLKTRIWLGTFFETAEDAARAYDEAARLMCGPRAR
TNFPYNPNAIPTSSSKLLSATLTAKLHKCYMASLQMTKQTQTQTQTARSQSADSDGVT
ANESHNLNRGVTETTEIKWEDGNANMQNFRPLEEDHIEQMIEELLHYGSIELCSVLPTQT
L*

>G1063 (241..966)
GTTAAAGAAGATGGATGGGCCACAAGTTGCTATATAAATCCTTCCACTTCTTGTGTATA
CTATTGCTTGAGTTCTGATTGGGCACAGTAGTACCATTTGCCATTCTCTCACACATACCG
TCTCTTTCTCTCATCATCAATCATCAATCATCCAAAAGAAAAACCCATAAATTTCACTT
GTAAGCTTTTACCAGTTTCTCTCCATACCCATTTTATCAGCTTCTCCATATCTTCTCT
ATGGATTCTGACATAATGAACATGATGATGCATCAGATGGAGAAGCTTCTGAGTTTGT
AACCCTAATTCCTCTTTCTCTCTCCCGACCACAACAACACTTACCCTTTTCTCTTAAAC
TCCACTCATTACCAGTCCGATCACTCAATGACCAACGAACAGGTTTCCGCTACGGTTCC
GGTTTACTCACTAACCTTCTTCTATCTCTCCCAACACAGCTTACTCTTCCGTTTTTCTT
GACAAAAGAAACAACAGTAACAACAACAATAATGGCACGAACTGGCAGCTATGCGAGAG
ATGATCTTCCGTATCGCCGTGATGCAACCGATCCATATCGATCCCGAGGCGTTAAGCCA

CCGAAGAGGAGGAACGTCAGGATCTCTAAAGATCCTCAAAGCGTGGCGGCTAGGCATAGA
AGGGAGAGAATAAGCGAGAGGATTTCGGATTTTGCAACGGCTTGTTCCTGGTGGGACGAAG
ATGGATACAGCTTCGATGCTCGATGAAGCAATTCATTATGTGAAGTTTTTAAAGAAACAG
GTGCAGTCTCTGGAGGAGCAGGCGGTGGTTACTGGCGGAGGGGGAGGAGGAGGAGGAAGG
GTTTTGATCGGTGGAGTGGGAATGACGGCGGCGAGTGGTGGTGGTGGCGGCGGGGAGTG
GTTATGAAAGGGTGTGGAACAGTGGGGACTCATCAGATGGTGGGCAATGCACAGATTCTT
AGATGATGATGATTTTTAATTTTATTATTATTATTAATGTTGGAGAAAAAGAGAAAAA
TGATTCTGGAGAGGGAAGCCAAGTAATTTATGTGAGAGTCTTTAATTTAACTTTATTTTC
TTGTTTAGATAATGTGTAATGATGGTTTTTAAAGCCAAAGACTCTCCATGGTTGTTGGAG
CGAGTTTG

>G1063 Amino Acid Sequence (domain in aa coordinates: 131-182)
MDSIDIMMMHMQEKLPEFCNPNSFFSPDHNNYPFLFNSTHYQSDHSMTNPEGFRYGS
GLLTNPSSISPNTAYSSVFLDKRNNNNNGTNMAAMREMI FRIAVMQPIHIDPEAVKP
PKRRNRVRSKDPQSVAAHRHRRERISERIRILQRLVPGGTKMDTASMLDEAIHYVKFLKKQ
VQSLLEQAVVTGGGGGGGGRVLIGGGGMTAASGGGGGGGVVMKCGTVGTHQMVGNAQIL
R*

>G1140 (67..729)

ATCCAAGATCCTCCAACCTCACAGAAAGGCAGATTCAAGAACAGTAGTGAAGGAGAGATCT
GGTAAATGGCGAGAGAGAAGATAAGGATAAAGAAGATTGATAACATAACAGCGAGACAA
GTTACTTTCTCAAAGAGAAGAAGAGGAATCTTCAAGAAAGCCGATGAACTTTCAGTTCTT
TGCGATGCTGATGTTGCTCTCATCATCTTCTCTGCCACCGGAAAGCTCTTCGAGTTCTCC
AGCTCAAGAATGAGAGACATATTGGGAAGGTATAGTCTTCATGCAAGTAACATCAACAAA
TTGATGGATCCACCTTCTACTCATCTCCGGCTTGAGAATTGTAACCTCTCCAGACTAAGT
AAGGAAGTCGAAGACAAAACCAAGCAGCTACGGAAACTGAGAGGAGAGGATCTTGATGGA
TTGAACCTAGAAGAGTTGCAGCGGCTGGAGAACTACTTGAATCCGGACTTAGCCGTGTG
TCTGAAAAGAAGGGCGAGTGTGTGATGAGCCAAATTTCTCACTTGAGAAACGGGGATCG
GAATTGGTGGATGAGAAATAAGAGACTGAGGGATAAACTAGAGACGTTGGAAGGGCAAAA
CTGACGACGCTTAAAGAGGCTTTGGAGACAGAGTCGGTGACCACAAATGTGTCAAGCTAC
GACAGTGGAACTCCCCTTGAGGATGACTCCGACACTTCCCTGAAGCTTGGGCTTCCATCT
TGGGAATGAATCTGAGAGAGAGAAAGATCCAGCAGAGTTGACTTCGATGGAAGCCACAA
ATATTAAGTCTACCTTTTCCCTTTCTTTCTTTGAATAAGTGTTGAAAAAGAATTGAGAT
GGGAAGGATGAATTCTCATTGCATTGCAGAGAAGCAAGTTTCAGATATTGTACGTGTTAT
TGGGTCTTTATAACTATTTTCTCCCCAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAA

>G1140 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)

MAREKIRIKKIDNITARQVTFKRRRGIFKKADELSVLCADVALIIFSATGKLFEPSSS
RMRDILGRYSLHASNINKLMDPPSTHLRLCENLSRLSKEVEDKTKQLRKLRGEDLDGLN
LEELQRLEKLLSEGLSRVSEKKGECVMSQIFSLEKRGSELVDENKRLRDKLETLEKRLT
TLKEALETESVTNVSSYDSGTPLEDDSDTSLKGLPSWE*

>G1425 (43..1005)

ACTCTCTCAAACCATAAAAAATATTCTCCGATCATCATTTTAATGGAGAGTACAGATTCT
TCCGGTGGTCTCCGCCGCCGCAACCAACCTCCCTCCAGGATTCCGGTTTCATCCAACA
GACGAAGAACCTGTAATTCATTACCTCAAACGCAAAGCAGATTCTGTTCTTTTACCAGTC
GCGATCATCGCCGACGTTGATCTTTACAAATTTGATCCATGGGAACCTCCCGCGAAAGCT
TCGTTTGGAGAACAAGAATGGTATTTTTTCAGTCCAAGAGATCGGAAATATCCCAACGGA
GCTAGACCTAACCAGCTGCGACTTCCGGTTATTGGAAAGCGACTGGTACAGATAAACCG
GTGATTTCAACCGCGCGTGGTGGTAGTAAAAAGTGGGAGTTAAAAAGGCTCTAGTGT
TACAGTGGTAAACCACCAAAAGGAGTTAAATCAGATTGGATTATGCATGAATATCGGTTA
ACTGATAATAAACCTACTCACATTTGTGACTTCGGCAACAAGAAAACTCTCTCAGGCTT
GATGATTGGGTGTTGTGTCGTATCTACAAGAAAAACAATAGTACAGCATCTAGACATCAT
CATCATCTTCATCATATTATCTAGATAATGATCATCATCGTCATGATATGATGATTGAT
GATGATCGATTCCGTCATGTTCTCTGGTCTTCACTTCCCGCGGATTTTTCTGACAAAT
AATGATCCGACGGCTATATATGATGGTGGCGGCGCGGATACGGAGGTGGAAGTTACTCG
ATGAATCATTTGTTTCGCATCTGGATCAAAGCAGGAGCAGTTGTTTCCACCGGTGATGATG
ATGACTAGTCTAAATCAAGATTCGGTATTGGATCGTCGTCGTCACCTAGCAAGAGATTT
AACGGCGGCGCGGCTTGGAGATTGTTGCACTTCTATGGCGGCGACGCCGTTAATGCAGAAC
CAAGGTGGGATTTACCAATTGCCTGGTTTGAATTGGTATTCTTGAAAAACAATTTACGATG
AAGAATTTTTAAATTTGTGTATATATATACGGTTTGAGTGATTAGGGGGCATTGGGGGA

TTTATTTACGGTTGATTATTATTGTAGTGTATAGAACTAAGGAGATTAAATTAAATAGA
TTGGAGGAAAAAAAAAAAAAAAAA
>G1425 Amino Acid Sequence (domain in AA coordinates: 20-173)
MESTDSSGGPPPPQPNLPPGFRFHPTDEELVIHYLKRKADSVPLPVAILIADVLYKFDPW
ELPAKASFGEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPVISTGGGSSKKVGV
KKALVFYSGKPPKGVKSDWIMHEYRLTDNKPHTICDFGNKKNSLRLDDWVLCRIYKKNS
TASRHHHHLHHIHLNDHHRHDMIDDRFRHVPPGLHFPALFSDNNDPTAIYDGGGGGY
GGGSYSMNHCFASGSKQEQLFPPVMMMTSLNQDSGIGSSSSPSKRFNGGGVGD CSTSMAA
TPLMQNQGGIYQLPGLNWYS*
>G1449 (105..581)
TAGACAGAGAGAAATAGAAATAGAGAGAGAGACATGAAGAGCACTCTCAATAGAGAAG
AGAAGGAAGCATGAAGCTAGCTCTGCAGCTTCAAGGTCTCATTAAAGGAGGTCTCTAACT
CTTGTTCTTTCATTTTCTTCATCCTCTGTCGACAGTACTAAACCTTCTCCTTCTGAATCTT
CTGTTAATCTCTCCCTTAGTCTCACATTCCTTCTACTTCTCCACAAAGAGAAGCAAGAC
AAGATTGGCCACCGATAAAGTCTAGATTAAAGAGATACACTAAAGGGTCGTCGTCTTCTTC
GTCGTGGTGATGACACTTCTCTCTTGTAAAGGTTTATATGGAAGGTGTCCCATTTGAA
GAAACTCGACCTTTGCGTATTCTCAGGCTACGAGAGTCTATTAGAAAATCTCTCTCACA
TGTTTCGATACTTCAATCATCTGCGGTAATCGAGATCGAAAACATCATGTTTGTACATATG
AAGACAAGGATGGAGATTGGATGATGGTCGGAGATATTCCATGGGATATGTTTCTTGAAA
CCGTGAGAAGACTAAAGATCAGAGACCGGAGAGGTATTAAACTTTGGATCGGTCAAGGC
TGTGATTGCGCAGTTACGAGACGTGTAAGATTTAGGCATTGATGAAGAGACTTGAGGCGG
GACGGAGCTATTGCTGCATATTGCAACAAAGGCCTTGAAGAAGTTGGAGAATTGATTGAT
GCATATATTTATTTATATGACACCTTTGAGTGTGTTTTTCTTATAAATAAATCACAATA
TCCAAGACTTCTCTTAA
>G1449 Amino Acid Sequence (domain in AA coordinates: 48-53,74-107,122-152)
MEVSNSSCSSSSSVSTKPSSESSVNLSSLTFTSTSPQREARQDWPIKSRRLDTLK
GRRLLRGGDDTSLFVKVMEGVPIGRKLDLCVFSGYESLLENLSHMFDTSIICGNRDRKH
HVLTYEDKGDWMMVGDIPWDMFLETVRRLKITRPERY*
>G1897 (1..678)
ATGCCTTCTGAATTCAGTGAATCTCGTCGGGTCTCTAAGATTCCCCACGGCCAAGGAGGA
TCTGTTGCGATTCCGACGGATCAACAAGAGCAGCTTCTTGCTCGCTGTGAATCAACC
AACACCAAGTTCTGTTACTACAACAATACTTCTCACAACCTCGTCATTTCTGCAAG
TCTTGTCGCGCTTACTGGACTCATGGAGGTACTCTCCGTGACATTCCCGTCGGTGGTGTT
TCCCGTAAAGCTCAAAACGTTCCCGGACTTATCTCTGCGCTACACCTCCGTTGTC
GGAAGCCGGAACTTCCCTTACAAGCTACGCCGTGTTCTTTTCCCTCAGTCGTCTTCCAAC
GGCGGTATCACGACGGCGAAGGGAAGTGCTTCGTCGTTCTATGGCGGTTTCAGCTCTTTG
ATCAACTACAACGCCCGGTGAGCAGAAATGGGCCTGGTGGCGGTTTAAATGGGCCAGAT
GCTTTTGGTCTTGGGCTTGGTCACGGGTCGTATTATGAGGACGTGAGATATGGGCAAGGA
ATAACGGTCTGGCCGTTTCAAGTGCGCTACTGATGCTGCAACTACTACAAGCCACATT
GCTCAAATAACCGCCACGTGGCAGTTTGAAGGTCAAGAGAGCAAAGTCGGGTTCTGTCT
GGAGACTACGTAGCGTGA
>G1897 Amino Acid Sequence (domain in AA coordinates:34-62)
MPSEFSERRVPKIPHGQGSVAIPTDQEQQLSCPRCESTNTKFCYNNYNFSQPRHFCK
SCRRYWITHGGTLRDIPVGGVSRKSSKRSRTYSSAATTSVVGSRNFPLQATPVLPQSSSN
GGITTAKGSASSFYGGFSSLINYNAAVSRNPGGGFNGPDAFGLGLGHGSYYEDVRYGQG
ITVWPFSSGATDAATTTSHIAQIPATWQFEGQESKVG FVSGDYVA*
>G2143 (89..784)
TCTTCTTCTCTCTCCATACCTTATCTCACCAGCTTCTCCATATCTCTCAAAGAAAAACA
AACCCTATAAATCCACAAAAAGGAGGATGGATAACTCCGACATTCTAATGAACATGAT
GATGCAGCAGATGGAGAAGCTTCTGAACACTTCTCTAACTCAAACCTAACCTAATCC
CCATAACATTATGATGCTTTCTGAATCCAACACCCACCGTTCTTCTCAACCCCACTCA
TTCTCATCTCCCATTTGACCAAACCATGCCTCACCACCAACCCGTTTAAATTTCCGGTA
CGCCCCCTCCCGTCATCATCTCTCCCGGAGAAGAGAGGAGGCTGCAGCGACAACGCCAA
CATGGCGGCGATGAGAGAGATGATCTTTGAATAGCCGTGATGCAGCCTATACATATTGA
TCCGGAATCCGTAAAGCCACCAAAGAGAAAGAACGTGAGGATCTCTAAGGATCCACAGAG
CGTGGCAGCTCGGCATCGAAGGAGAGGATAAGCGAGCGGATTCGGATTCTTCAGCGGCT
TGTTCCCGGTGGGACTAAGATGGATACGGCGTCGATGCTCGATGAGGCTATCCATTACGT

TAAGTTTCTCAAGAAGCAAGTGCAGTCGCTGGAGGAACATGCGGTGGTTAACGGCGGAGG
AATGACGGCGGTGGCCGGAGGAGCACTTGCGGGTACTGTTGGTGGAGGATATGGAGGAAA
AGGGTGTGGCATTATGCGGTCTGATCATCACCAGATGCTTGGAAATGCACAGATTCTTAG
ATGATGATGATGTTGATTTTTAAATATATATCATATGTTTATTAATATGACGGGAAAAAA
TATTATCGAGGGAGTTGAATTTAGTATCATGAACTATGAGAGCATTTTTTTTAAATGTT
TTTATCTTTCCGGGTTTCGATAATGTTTGGGATGGTTAATTAACAATTTAAAAGTCAGAC
AAGTTGGTTGTAAAGACTAAAGAATAAGCATAGTTTATCAATTTATCATTACTAAATGAA
ATAG

>G2143 Amino Acid Sequence (domain in aa coordinates: 128-179)

MDNSDILMNMMMQMEKLPFHFSNSNPNNPHNIMMLSESNTHPFFFNPTHSHLPFDQTM
PHHQPLNFRYAPSPSSSLPEKRGCCSDNANMAAMREMI FRIAVMQPIHIDPESVKPKR
KNVRISKDPQSVAAHRHRRERISERIRILQRLVPGGTKMDTASMLDEAIHYVKFLKKQVQS
LEEHAVVNGGGMTAVAGGALAGTVGGGYGGKGCIMRSDHHQMLGNAQILR*

>G2535 (1..1005)

ATGAACATATCAGTAAACGGACAGTCACAAGTACCTCCTGGCTTTAGGTTTACCCCAACC
GAGGAAGAGCTCTTGAAGTATTACCTCCGCAAGAAAATCTCTAACATCAAGATCGATCTC
GATGTTATTCTGACATTGATCTCAACAAGCTCGAGCCTTGGGATATTCAAGAGATGTGT
AAGATTGGAACGACGCGCAAAACGATTGGTACTTTTATAGCCATAAGGACAAGAAGTAT
CCCACCGGGACTAGAACCAACAGAGCCACCACGGTCGGATTTTGGAAAGCGACGGGACGT
GACAAGACCATATATACCAATGGTGATAGAATCGGGATGCGAAAGACGCTTGTCTTCTAC
AAAGGTCGAGCCCTCATGGTCAGAAATCCGATTGGATCATGCACGAATATAGACTCGAC
GAGAGTGATTAAATCTCCTCGTGTGGCGATCATGACGTCAACGTAGAAACGTGTGATGTC
ATAGGAAGTGACGAAGGATGGGTGGTGTGTCGTGTTTTCAAGAAAAATAACCTTTGCAAA
AACATGATTAGTAGTACCCCGGAGTTCGGTGAAAACGCCGTCGTTCAATGAGGAGACT
ATCGAGCAACTTCTCGAAGTTATGGGGCAATCTTGTAAGGAGAGATAGTTTTAGACCCT
TTCTTAAAACTCCCTAACCTCGAATGCCATAACAACACCACCATCACGAGTTATCAGTGG
TTAATCGACGACCAAGTCAACAACCTGCCACGTGAGCAAGTTATGGATCCCAGCTTCATC
ACTAGCTGGGCGCTTTGGATCGGCTCGTTGCCCTCACAGTTAAATGGGCCCCAAGTCGTAT
TCAATACCAGCCGTTAATGAGACTTCACAATCACCCTATCATGGACTGAACCGGTCCGGT
TGTAATACCGGTTTAAACACCAGATTACTATATACCGGAGATTGATTTATGGAACGAGGCA
GATTTCCGCGAGAACGACATGCCACTTGTTGAACGGTAGTGGATAA

>G2535 Amino Acid Sequence (conserved domain in AA coordinates: 11-114)

MNISVNGSQVPPGFRFHPTEELLKYYLRKKISNIKIDLDVIPDIDLNKLEPWDIQEMC
KIGTTPQNDWYFYSHKDKKYPTRTNRATTVGFWKATGRDKTIYTNDRIGMRKTLVYF
KGRAPHGQKSDWIMHEYRLDESVLISSCGDHVDNVETCDVIGSDEGWVVCRVFKNNLCK
NMISSSPASSVKTPSFNEETIEQLLEVGMQSKGEIVLDPFLKLPNLECHNNTTITSYQW
LIDDQVNNCHVSKVMDPSFITSWAALDRLVASQLNGPNSYSIPAVNETSQSPYHGLNRS
CNTGLTPDYIPEIDLWNEADPARTTCHLLNGSG*

>G2557 (94..1215)

TCGACTTCCTGTGAACATCATCTGTTTGTCTCTTCTCCGGTTTCACTTTTTCATGTCCT
GCCGTTATTACAACGAGGATTGTGTTTGATCCGATGGAAGGATTGGAATCTGTGTACGCT
CAAGCTATGTATGGAATGACACGAGAGAGCAAAATCATGGAGCATCAAGGATCAGATTTG
ATTTGGGGAGGAAATGAGCTAATGGCTCGAGAACTCTGTCTTCTTCTTCTTATCACCAC
CAACTCATTAAATCCGAATCTTAGCAGCTGTTTCATGTCTGATCTTGGAGTCTTAGGTGAG
ATTCAACAGCAGCAACATGTTGGCAACAGAGCTAGCTCGATAGATCCATCATCACTCGAT
TGTTTGTATTCTGCGACGTCGAATAGCAACAACACCTCGACGGAGGACGATGAAGGAATA
TCTGTGCTTTTCTCAGATTGTCAGACTCTTTGGAGCTTGGTGGAGTCTCATCTGCAGAG
TCTGAGAACAGAGAGATCACTACTGAGACGACAACAACGATAAAGCCTAAGCCTTTGAAG
AGAAACAGAGGAGGAGATGGAGGAAGTACTGAGACTACAACAACAACAACAAACCTAAG
TCTTTGAAGAGAAACAGAGGAGACGAGACAGGAAGTCACTTTAGTCTTGTTCATCCTCAA
GATGATTCGGAGAAAGGAGGTTTCAAGCTTATATACGATGAGAATCAATCGAAATCAAAG
AAACCAAGAACAGAGAAAGAACGAGGCGGTTCTTGAACATTAGTTTCCAACATTCAACT
TGTTTGTCTGACAATGTGCGAGCCCGATGCTGAGGCGATTGCACAAATGAAGGAGATGATA
TACAGAGCGGCTGCATTTAGACCGGTGAATTTCCGGTTAGAGATTGTGGAGAAGCCTAAG
AGGAAGAACGTCAAGATATCGACGGATCCTCAAACGGTTGCAGCGAGACAGAGAAGGGAG
AGGATAAGTGAGAAGATTAGGGTTTACAAAACATTGGTTCCAGGTGGGACGAAGATGGAT
ACTGCATCAATGCTTGATGAAGCTGCTAATTATCTCAAGTTCTTAGAGCACAAGTAAAA

GCTTTAGAAAACCTTGAGACCCAAGCTTGACCAAACCAATCTCTCTTTCTCTCTGCTCCT
 ACATCGTTTCCATTATTTCCACCCATCTTTCTTCCATTGCAAAATCCTAATCAAATCCAT
 CATCCAGAGTGTGACAGATTATAAACTTTTGAGTTTCATCATCATCAACAGAATCATGG
 CGTCTTGATTGTTTTAGCAGTTCTCAAGAAAGGCAACTTCTGTGACAAGGGTGGTGTCTGG
 GCAGTGTGTTTACACTTTCCAGTCTTTGTTTTGCATTTCTTTTATATAAAGTTTGAT
 TTTATATAGAATCTGTGGAATTCGAGGGTTGAAATATTGTGAAAAACAGAGCCGCAAGAG
 GTTAATTACAGTCTCTGCAATATTTTCAACCTTTTATTACTTTATTAGAGTAAAGATAGC
 GT

>G2557 Amino Acid Sequence (domain in aa coordinates: 278-328)

MEGLESVYAQAMYGMTRESKIMEHQSDLIWGGNELMARELCSSSSYHHQLINPNLSSCF
 MSDLGVLGEIQQQHVGNRASSIDPSSLDCLLSATSNNTSTEDDEGISVLPSCQTLW
 SFGGVSSAESENREITTTETTTTIKPKPLKRNRRGGDGGTTETTTTTPKPSLKRNRRDETG
 SHFSLVHPQDDSEKGGFKLIYDENQSKSKPRTEKERGGSSNISFQHSTCLSDNVEPDAE
 AIAQMKEMIYRAAARFVNFGLIVEKPKRKNVKISTDPQTVAARQRRERISEKIRVLQT
 LVPGGTAKMDTASMLDEAANYLKFLRAQVKALENLRPKLDQTNLSFSSAPTSFPLFHPSPFL
 PLQNPNIHHEPC*

>G259 (52..786)

GAGATCTTCTACTACTGTGTTTTCTTCAAGAATAATAATTTTCGTTTATATATGGAAGAT
 GCTGGTGAACATTTACGGTGTAAACGATAACGTTAACGACGAGGAGCGTTTGCCATTGGAG
 TTTATGATCGGAACTCAACATCCACGGCGGAGCTACAGCCGCCTCCACCGTTCTTGGTA
 AAGACATACAAAGTGGTGGAGGATCCGACGACGGACGGGGTTATATCTTGGAAACGAATAC
 GGAAGTGGTTTCGTCGTGTGGCAGCCGCGAGAATTCGCTAGAGATCTGTTACCAACACTT
 TTCAAGCATTGCAACTTCTCTAGCTTCGTTCCGCGACTCAATACTTACGGTTTTCGAAAA
 GTAACGACGATAAGATGGGAATTTAGTAATGAGATGTTTCGAAAGGGGCAAAGAGAGCTT
 ATGAGCAATATCCGAAGAAGGAAGAGCCAAACATTGGTCACACAACAAGTCTAATCACCAG
 GTTGTACCAACAACAACGATGGTGAATCAAGAAGGTCATCAACGGATTGGGATTGATCAT
 CACCATGAGGATCAACAGTCTTCCGCCACTTCATCCTCTTTCGTATACACTGCATTACTC
 GACGAAAAACAATGCTTGAAGAATGAAAACGAGTTATTAAGCTGCGAACTTGGGAAAAAC
 AAGAAGAAATGCAAGCAGCTTATGGAGTTGGTGGAGAGATACAGAGGAGAAGACGAAGAT
 GCAACTGATGAAAGTGAATGATGAAGAAGATGAAGGGCTTAAGTTGTTCCGAGTAAACTT
 GAATGAACTAGATTGCTAGATTGATATTCGTAATATACCAGTTTCTTCATATTCTTAGA
 AGTTTTGCATAACTATATATAGTACTCTTTAAGACATGCAAGATCAGAACATATG

>G259 Amino Acid Sequence (domain in AA coordinates: 27-131)

MEDAGEHLRCNDNVNDEERLPLEFMIGNSTSTAELOPPPPFLVKTYKVVEDPTTDGVISW
 NEYGTGFVWQPAEFARDLLPTLFKHCNFSFVRQLNTYGRFVTTIRWEFSNEMFRKQ
 RELMSNIRRRKSQHWSHNSNHQVVPVPTTMVNVQEGHQIRIGIDHHHEDQQSSATSSSFVYT
 ALLDENKCLKNENELLSCELBKTKKKCKQLMELVERYRGEDEDATDESDDDEDEGLKLF
 VKLE*

>G353 (82..570)

ACCAAACCTCAAAAAACACAAACCACAAGAGGATCATTTTCATTTTTTATTGTTTCGTTTTA
 ATCATCATCATCAGAAGAAAAATGGTTGCGATATCGGAGATCAAGTCGACGGTGGATGTC
 ACGGCGGCGAATTTGTTGATGCTTTTATCTAGAGTTGGACAAAGAAACGTTGACGGTGGC
 GATCAAAAACGCGTTTTTACATGTAAAACGTGTTTGAAGCAGTTTCATTCGTTCCAAGCC
 TTAGGAGGTACCGTGCGAGTCACAAGAAGCCTAACAAACGACGCTTTGTCGTCTGGATTG
 ATGAAGAAGGTGAAAACGTCGTCGATCCTTGTCCCATATGTGGAGTGGAGTTTCCGATG
 GGACAAGCTTTGGGAGGACACATGAGGAGACACAGGAACGAGAGTGGGGCTGCTGGTGGC
 GCGTTGGTTACACGCGCTTTGTTGCCGGAGCCACGGTGACTACGTTGAAGAAATCTAGC
 AGTGGGAAGAGAGTGGCTTGTGATCTGAGTCTAGGGATGGTGGACAATTGAATCTC
 AAGTTGGAGCTTGAAGAAGCAGTTTATTGATTTTATTTATTTTCTTAAATTTTCTGAAT
 ATATTGTTTTCTCTCATTCTTTGAATTTTCTTAATATTCTAGATTATACATACATCCGC
 AGATTAGGAAACCTTCATAGAGTGAATCTTTCTTTCTGTAAAAATATATTTTACTTG
 TAGCAAA

>G353 Amino Acid Sequence (domain in aa coordinates: 41-61, 84-104)

MVAISEIKSTVDVTAANCLMLLSRVQENVDGGDQKRVFTCKTCLKQFHSFQALGGHRS
 HKKPNNDALSSGLMKVKVTSHPICGVEFPMGQALGGHMRRHRNESGAAGGALVTRAL
 LPEPTVITLKKSSSGKRVAACLDLSLGMVDNLNLKLELGRTVY*

>G354 (27..533)

CCTAGAAGTCACTAAGTCGATTCAAAATGGTTGCGAGAAGTGAGGAAATTGTGATAGTGG
AAGAAGATACGACTGCGAAATGTTTGATGTTGTTATCAAGAGTCGGAGAATGCGGCGGCG
GCTCGGGGGGAGATGAACGTGTTTTCCGATGCAAGACTTGTCTTAAAGAGTTCTCATCGT
TTCAAGCTTTGGGAGGTTCATCGTGCAAGCCACAAGAACTTATCAACAGTGACAATCCAT
CACTTCTTGGATCCTTGTCCAACAAGAACTAAAACGTCTCATCCTTGTCCGATATGTG
GAGTGAAGTTTCCGATGGGACAGCTCTTGGTGGTCACATGAGGAGACATAGGAACGAGA
AAGTCTCAGGCTCGTTGGTTACACGTTCTTTTCTACCGGAGACGACGACGCGTGACGGCTT
TGAAGAAATTTAGTAGTGGGAAGAGAGTGGCTTGTGTTGGATTGGACTTAGATTCGATGG
AGAGTTTGGTCAATTGGAAGTTGGAGTTGGGAAGAACGATTCTTGGAGTTAAGTTTTTG
GGTTGTATACAGTTTACATGATTTTGTATCTTTGTTGATCCAATTATCGTACCGATCG
ATGTGAATATTATTTTGATACAATAAAA

>G354 Amino Acid Sequence (domain in aa coordinates: 42-62, 88-109)

MVARSEEIVVEEDTTAKCLMLLSRVGECGGGCGDERVFRCKTCLKEFSSFQALGGHRA
SHKKLINSNDNPSLLGSLSNKKTSTHPCPICGVKPFMGQALGGHMRHRNEKVSGLVTR
SFLPETTTVTALKKFSSGKRVACLDLDDLSMESLVNWKLELGRITISWS*

>G638 (86..1861)

GAATTAAGAGTTTAAACCTTTACCTTTTTTCCCTTCACTATCGATAATTGATCTTCTCT
TTCGGCTGAATATAAATCTGAAAAATGGATCAAGATCAGCATCCTCAGTACGGTATACC
GGAGCTCCGGCAGCTCATGAAAGGCGGAGGAAGGACGACTACTACAACACCGTCTACTTC
TTCTCATTTTCCCTCTGATTTCTTTCGGTTTAAACCTTGCTCCGGTGCAGCCACCGCCACA
CCGTCTTCATCAGTTCACTACTGATCAAGATATGGGTTTCTTGCCACGTGGCATAACATGG
ATTGGGTGGAGGTTCTTCAACGGCTGGAAATAACAGTAACTTAAACGCGAGTACTAGTGG
TGGAGGAGTTGGGTTTAGTGGGTTTCTTGACGGTGGTGGTTTCCGGCAGCGGAGTAGGAGG
AGACGGTGGAGGAACGGAAGGTGGCCGAGACAAGAAACCCCTAACTCTGTTGGAATTAG
ATCTCGTCTTGATCATAAATTCAAAGAAGCTAATCATAAAGGACCTCTTGGGATGAAGT
TTCTAGGATTATGTCCGAGGAACATGGATACCAAGGAGTGGGAAGAAATGCAGAGAGAA
GTTTGAGAATCTGTACAAATACTATAGTAAGACTAAAGAAGGCGAAGCCGGAAGACAAGA
CGGAAACATCACAGATTTTTCCGCCAGCTCCAAGCGCTATACGGGGATTCTAATAACTT
GGTTTCTTGTTCCCAATCATAACACGCGAGTTTCATGAGCAGTGCTCTTCATGGTTTCCATAC
TCAAAACCCCTATGAACGTTGCTACAAACAGTCCAACATCCATAACGTTGATAGTGTCA
TGTTTTTCATCAAAGCCTTAGTCTTTCTAACAACTACAACCTCCTCCGAGCTTGAGCTGAT
GACTTCTCTTTCGGAAGGGAATGATTCTAGTAGTAGAAGGAAAAAGAGGAGTTGGAAGC
GAAGATAAAGGAGTTCAATTGATACGAACATGAAAAGGTTGATAGAGAGGCAAGATGTTTG
GCTTGAGAAGTTGACAAAGGTTATGAAGACAAAGAGGAACAACGGATGATGAAAGAAGA
GGAATGGAGGAAGATTGAAGCTGCAAGGATTGATAAAGAGCATTGTTTTGGGCTAAAGA
GAGGCGAGGATGGAAGCTAGGGATGTTGCGGTGATTGAGGCAATTGCAATATTGACAGG
AAAGCCATTGATAAAGCCGCTGTGTTTCATCCCGGAAGAGAGGACAAATGGTAATAATGA
GATCCGAAACAATAGTGAGACACAGAATGAGAATGGAAGCGATCAAACGATGACTAACAA
TGTTTGTGTTAAAGGAAGTAGTAGCTGCTGGGCTGAGCAAGAGATTTTAAAGCTTATGGA
GATAAGAACGAGCATGGACTCGACCTTTCAAGAGATATTAGGAGGGTGCTCGGATGAGTT
TCTATGGGAGGAAATCGCAGCGAAGTTGATTAGTTAGGGTTTGATCAGAGAAGTGCCTT
ATTATGCAAGGAAAAGTGGGAATGGATAAGCAATGGAATGAGGAAAGAAAAGAAGCAAAT
CAACAAGAAAAGAAAGGATAATTCGTCCAGCTGCGGCGGTGACTACCCGAGAAACGAAGA
AAATCCAATCTACAATAATCGAGAAAGTGGATATAATGATAATGATCCGCATCAAATCAA
CGAACAAGGCAATGTAGGTTCTTCAACATCAAACGCAAACGCAAACGCAAACGTAACCAC
TGGAAATCCGAGCGGTGCAATGGCTGCTAGTACAAACTGCTTCCCGTTCTTCATGGGAGA
TGGAGATCAGAATTTGTGGGAGAGTTATGCTTTGAGGCTCAGTAAAGAAGAGAATCAGTA
AGTAATTTCTCTTAATGAAGAAGAAGAAGTAATCATGTGGTTAACTAATTCTTTTGGT
TAGCTATATATGAGATAAACCTTGACTTAGCTATTATATGTACATGTGCTTAGAATTA
AGAAATATTGTTGGGGCTTAACGAATTATATATCAGCATATATAAGATGAGAGTCTAAG
AATTATATCAAATTAGGCTTTAACCACGTACGATTATATATTATGTTTTTCATGTATTTA
TTCTGTAAGACTTTTTAATATCAATCTTCTCTAAA

>G638 Amino Acid Sequence (domain in AA coordinates: 119-206)

MDQDQHPQYGIPELRQLMKGGGRTTTTTPTSSHFPSDFFGFNLPVQPPPHRLHQFTTD
QDMGFLPRGIHGLGGSSSTAGNNSNLNASTSGGGVGFSGFLDGGGFGSGVGGDGGGTGRW
PRQETLTLLLEIRSRDLHKFKKANHKGPLWDEVSRIMSEEHGYQSRGKKCREKFENLYKYY
SKTKEGEAGRQDGKHHRRFRQLQALYGDSSNNLVSCPNNHTQFMSSALHGFHTQNPMMNVAT

TTSNIHNVDSVHGFHQSLSLSNYNSSSELELMTSSSEGNDSRRKKRSWKAKIKEFIDT
 NMKRLIERQDVWLEKLTKEVIEDKEEQRMKEEWRKIEAARIDKEHLFWAKERARMEARD
 VAVIEALQYLTKGLIKPLKSSPEERTNGNNEIRNNSETQNEGSDQTMNNVCVKGSSS
 CWGEQEILKLMEIRTSMDSTFQEILGGCSDEFLWEEIAAKLIQLGFDQRSALLCKEKWEW
 ISNGMRKEKKQINKKRKDNSSSCGVYPRNEENPIYNNRESGYNDNDPHQINEQGNVGSS
 TSNANANANVTGPNPSGAMAASTNCFPFPMGDGDQNLWESYGLRLSKEENQ*

>G869 (428..1402)

AGGAACAGTGAAAGGTTTCGGTTTTTTGGGTTTCGATCTGATAATCAACAAGAAAAAGGG
 TTTGATTTATGTCGGCTGGGTTTGAATCGACTGTGATTTTGTCTTTGATTTCATATCTCTT
 CTCCGATTTTCATCATCTCTTCCCATCATCGTCGTTTGAATCTTGTCTTCTCAACG
 CTCTTCACTTCTGCTGTGAAGCAAGGAGGCTTGTCTGGAGACTCCTTCTCTTTCATGC
 GCTTAAGACCCAAAAGGACTTGTCTAGTGTGAAGTCTTGGGGGTTTTACATAAAGC
 AGCAAAAGTTTTCTTTTTTCATAGTTCGCTGAGAGTTTGAAGTTTGATACCAAAAAAGT
 TTTGACCTTTTAGAGTGATTTTTTGTCTTCTGTCTTCTGGGTATTTTTGAGGAGTGGG
 TTTAACAATGGTTGCGATTAGAAAGGAACAGTCTTTGAGTGGTGTAGTAGCGAGATTAA
 GAAGAGAGCTAAGAGAAACACTCTATCGTCCCTTCTCAAGAAACCCAACTTTGAGGAA
 AGTCCGTATTATTGTGAATGATCCTTATGCTACTGATGATTCCTCTAGTGATGAGGAAGA
 GCTTAAGGTTCTTCAAGCAAGGAAATGAACAGTATCGTTCTGAGATTAACTTTCTCTTC
 TATGGAAGTTTTCTGAACAGCCTTCTGAGAGTCTTCTCAGGACAGTACTAAAACGTATGG
 CAAGATAGCTGTGTGCTGCTTCTCCTGCTGTTCTTAGGAAGAAGCCTGTTGGTGTTAGGCA
 AAGGAAATGGGGGAAATGGGCTGCTGAGATTAGAGATCCTATTAAGAAACTAGGACTTG
 GTTGGGTACTTTTTGATACTCTTGAAGAAGCTGCTAAAGCTTATGATGCTAAGAAGCTTGA
 GTTTGATGCTATTGTTGCTGGAATGTGTCCACTACTAAACGTGATGTTTCTTCATCTGA
 GACTAGCCAATGCTCTCGTTCTTCACCTGTTGTTCTGTTGAGCAAGATGACACTTCTGC
 ATCAGTCTCTCATTGTGTCAACAACCTGATGACGTCFCGACCGTTGCTCCAACGTCTCC
 AACTCCAAATGTTCTGCTGGTGGAAACAAGGAAACGTTGTTGATTTGACTTTACTAA
 TCTACAGATCCCTGATTTTGGTTTTCTTGGCAGAGGAGCAACAAGACCTAGACTTCGATTG
 TTTCTCGCGGATGATCAGTTTGTGATTTTCGGCTTGCTTGATGACATTCAAGGATTCTGA
 AGATAACGGTCCAAGTGCGTTTACCAGATTTCGACTTTGCGGATGTTGAAGATCTTCAGCT
 AGCTGACTCTAGTTTCGGTTTTCTTGATCAACTTGCTCCTATCAACATCTCTTGCCCAT
 AAAAGTTTTCAGCTTCATAGGATCTTGCTTAGTAATGTTAAGTGAGAAGAGTGTTTTG
 TTTTTTCGTTTATGCTTTAGTAATTTAAGACATACAAAAGTGTGTGTTCCGGATTGTAGT
 AAGATCTTAAGACATAAAGCCGGGTTTTGCAATTAGGAATCGAGTTTAAATGAAGTTTA
 GTTTATGTTTG

>G869 Amino Acid Sequence (domain in AA coordinates: 109-177)

MVAIRKEQSLSGVSSEIKKRAKRNLTSSLPQETQPLRKVRIIVNDPYATDDSSSDDEELK
 VPKPRKMKRIVREINFPSMEVSEQPSSESSQDSTKTDGKIAVSASPAVPRKKPVGVRQRK
 WGWAAEIRDPIKTRTWLGTFTDLEAAKAYDAKKLEFDAIVAGNVSTTKRDVSSSETS
 QCSRSSPVVPVEQDDTASALTCVNNPDDVSTVAPTPTPNVPAGGNKETLFDFTNLQ
 IPDFGFLAEEQQDLDFDCFLADDQFDDFGLLDDIQGFEDNGPSALPDFDFADVEDLQLAD
 SSFGFLDQLAPINISCLPKSFAAS*

>G1645 (25..1104)

CGTCGACCTCCCAACACTAACTCCATGTTTATAACGGAAAAACAAGTGTGGATGGATGAG
 ATCGTCGCAAGAAGAGCTTCTTCTTCTTGGGACTTCCCTTTCAACGACATTAATATTCAT
 CAGCATCATCATCGTCACTGCAACACAAGTCATGAGTTTGAATCTTGAAGAGTCTCTT
 GGAGATGTAGCGGTTACGAAGAAGAGAGTAATAATAATAACCTAATTTTCAGTAACAGC
 GAGAGTGGTAAGAAGGAGACAACAGATAGTGGTCACTTGGTCTCTCGTCGTTCAAAA
 CCATCGGTCTTGGGGAGAGGACATTGGAGACCAGCTGAAGATGTTAAACTCAAAGAGCTT
 GTCTCCATTTACGGCCCAAAAACCTGGAACCTCATAGCTGAAAAGCTTCAAGGAAGATCT
 GGGAGAGCTGTAGACTACGATGGTTTAACCAATTGGACCCGAGGATAAACCGAAGAGCT
 TTCACAGAAGAAGAAGAGGAGAGGCTGATGCAAGCACATAGGCTTTATGGTAACAAATGG
 GCAATGATTGCGAGGCTTTTCCCTGGTAGAACTGATAATTCACTGAAGAACCATTGGCAT
 GTTGTCTAGGCTCGTAAGTATAGAGAACACTCTTCTGCTTACCGTAGGAGAAAGCTTATG
 AGTAATAATCCACTTAAACCTCACCTCACCAATAATCATCATCCTAACCTAACCTAAT
 TACCACTCTTTTATCTCCACTAATCATTACTTCGCTCAGCCTTTCCCCGAGTTTAATTTG
 ACTCATCACTGGTTAATAATGCCCTATCACGAGTGACCATAACCAGCTTGTGTTGCCT
 TTCCATTGCTTTCAAGTTATGAGAACAAATGAACCTCCGATGGTTGTGAGTATGTTGGC

AACCAAATGATGGTCGGCGATAACGTTGGTGCCACGTCAGACGCGTTATGCAATATTCCG
CACATTGACCCTAGTAACCAAGAGAAACCGGAGCCAAATGATGCAATGCATTGGATCGGA
ATGGACGCGGTAGATGAGGAGGTGTTGAAAAGGCTAAGCAGCAACCACATTTTTTCGAT
TTTCTTGGCTTGGGGACGGCGTGAATGTTGAACAAATTGGTGTTAATCAGATAACGACAG
TGGC

>G1645 Amino Acid Sequence (domain in AA coordinates: 90-210)

MFITEKQVWMDEIVARRASSSWDFPNDINIHQHHHRHCNTSHEFEILKSPLGDVAVHEE
ESNNNNPNFNSNESGKKETDSDGQSWSSSSSKPSVLGRGHRPAEDVKLKLVSIIYGPQN
WNLIAEKLQGRSGKSCRLRWFNQLDPRINRRAFTEEEERLMQAHRLYGNKWAMIARLFP
GRTDNSVKNHWHVVMARKYREHSSAYRRRKLMSNNPLKPHLTNNHHPNPNPNYHSFISTN
HYFAQPFPEFNLTHLVLNAPITSDHNQLVLPFHCFQGYENNEPPMVVSMFGNQMMVGDN
VGATSDALCNIPHIDPSNQEKPEPNDAMHWIGMDAVIDEEVFEKAKQQPHFFDFLGLGTA*

>G1038 (240..1574)

GCTCGTTTTCAAATTAAAAACAGGGAGAAATTTGGAAATTCAGTACGACGGGAGATAAA
ACCTAACATACGCCATGGTGACCGTTATCTAAACTACGCCAAAATATTTGAAGTGTCGTC
GTTTCATAATAAAACGCAACAAAAACCCACTCCCACTTTCTCCTTTCCAAAAAAGAAC
TCTCGCCACTTTCTCTGCTCTTTCTTTCTCTCTCTCTTTCTTGTTTTCGCCGGCGATCA
TGGAGAAAAGCGGCTTCTCTCCGTCGGTCTAAGGGTTCTTGTCGTAGACGATGATCCAA
CTTGGCTCAAGATTCTCGAGAAAATGCTCAAGAAGTGTCTTACGAAGTAACGACCTGTG
GATTAGCTAGAGAGGCTTTGAGGTTGCTGAGGGAGCGTAAAGATGGATATGATATCGTGA
TCAGCGATGTGAACATGCCTGACATGGATGGTTTCAAGCTTCTTGAGCATGTTGGTCTTG
AATTAGACCTCCCTGTAATAATGATGTCGGTGGACGGCGAAACAAGCCGAGTGATGAAGG
GAGTGCACACGGGAGCTTGTGATTACCTCTGAAGCCGATAAGAATGAAGGAGTTAAAGA
TTATATGGCAACATGTTCTGAGAAAAGAGCTTCAAGAAGTGAGAGATATCGAAGGCTGTG
GATACGAAGGAGGAGCGGATTGGATCACTCGATACGATGAAGCACATTTTCTTGAGGTTG
GTGAAGATGTTTTCTTTGGGAAAAGAGAAAAGACTTTGACTTTGAGAAGAAGCTTCTTC
AAGATGAGAGTGATGTCATCATCTTCTTCTTCAAGAAAGCTAGAGTTGTTTGGTCTTTTG
AGCTTCATCATAAGTTTGTCAACGCCGTTAACCAAATCGGATGCGATCACAAAGCTGGTC
CCAAGAAGATATTGGATCTCATGAATGTTCCATGGCTCACTAGAGAAAATGTTGCAAGCC
ACCTTCAGAAATATAGACTTTACCTGAGCAGATTAGAGAAAGGAAAGGAGCTCAAGTGTT
ATTCAGGTGGCGTGAAGAATGCGGATTCTCTCCAAAGATGTCGAAGTGAATTCAGGCT
ACCAAGCCCTGGGAGGAGCAGCTATGTATTCTCTGGAGGAAATCTCTGATCCAAAAAG
TGATCATGCCTCCGAAAAACAAAAGACGCGTATAGGATTGATCCTCCCATTTCTCTCT
CTGCGTTTGACTCTCTGCTTCTTGAATGATGTTCCAGAGGTCCTGAATCGAAGCCGG
TTCTGTATGAGAATAGCTTTCTCCAGCAACAACCATGCGCAAGTCAAAGTTCTTATGTTG
CAATTTCTGCACCATCTCTCATGGAGGAGGAAATGAAGCCTCCTTATGAGACACCAGCAG
GAGGCAGTAGTGTGAATGCAGATGAGTTTCTCATGCCACAAGACAAGATCCCTACTGTAA
CCCTTCAAGATTGGATCCCTCTGCCATGAAGCTGCAGGAGTTCAACACAGAAGGCGATT
CTGAAGAAGCTTGAAGTGGGAACTTCCAGAATCACATCATTCTGTTTCTTTAGACACTG
ACTTAGACTTGACTTGGCTTCAAGGCGAGCGTTTCTTGCAAACACCGACTCCAGTTTCAA
GATACAGTAGTAGCCCATCACTCCTATCTGAGCTCCCAGCCCACCTTAATTGGTATGGAA
ATGAGCGGCTGCCTGACCCTGACGAGTATTCCTTCATGGTAGACCAAGGTTTATTCTAT
CTTAACCTTGTTTCAATAACTTCTTTTCGTATATTGGTTGGTGTAAATGCAGAAAGATTTT
GTGGGTATACCTGAAAATAATCTTGCTTTCCCAAGAACCTTCCATGATCGGATGCATTGT
ACAATAATCCAGAGTGTGCTAGGCTAATTACACCAACAGGTTGATGACAGTGATAAGG
CCACATGTTTACACCCGTCGCTTAAGATCTTTACTGTCACTGGAAGGAAA

>G1038 Amino Acid Sequence (domain in AA coordinates: 198-247)

MEKSGFSPVGLRVLVDDPTWLKILEKMLKKCSYEVTTCLAREALRLLRERKDGDIV
ISDVNMPDMDGFKLLEHVGLLELDLPVIMMSVDGETSRVMKGVHTGACDYLLKPIRMKELK
IIWQHVLRLKKLQEVRIEGCGYEGGADWITRYDEAHFLGGGEDVSFGKKRKDFDFEKKLL
QDESDPSSSSSKARVVWSFELHHKFVNNAVNIQCDHKAGPKKILDLMNVPLWTRENVAS
HLQKYRLYLSRLEKKGKELKCYSGGVKNADSSPKDVEVNSGYQSPGRSSYVFGGNSLIQK
ATEIDPKPLASASLSDPNTDVIMPPKTKKTRIGFDPPISSAFDSLPLPNDVPEVLESKP
VLYENSFLQQQLPLPSQSSYVAISAPSLMEEEMKPPYETPAGGSSVNADEFMLPQDKIPTV
TLQDLDPAMKLQEFNTEGDSEEA*

>G1073 (62..874)

[illegible]

cctaaaggcaagaggttagagcttctgctggcaatattacctgataacaacgggttcacttt
 atggtgatcttaagagaatctgtgaaaccgagcttggttgatatctcaatggtgtctca
 caaaacatgtgttcaagatttagcaaacagtatctggcagatgtatcccttaaaatcaacg
 taaagatgggaggaaggaacacagttctagtagacgccataagctgtagaattccactgg
 ttagcgatataccgacaatcatttttggcgcagacgtgactcaccagagaacggggaag
 agtcaagcccttcaatcgctgctgtgtgtgcttctcaagactggcctgaagtgacaaaat
 atgcggttttagtttgctcaagctcacaggcaagaacttatacaagatttgataaaa
 catggcaagatcctgttcgcggtactgttagtgcggtatgatcaggaccttcttatct
 catttagaaaagcaacagggcaaaaaccgcttcgaattatcttttatcgtgatggagtaa
 gcgaagggaattctatcaagttttactctatgagttggatgcaattcgaaaggcttggtg
 catcgcttgaaccgaattatcagccaccggtgacattcatagttgtacagaagcgtcacc
 acactcggttgggttctaataatcaccgagacaaaaacagtagtaccgaagcggaata
 tcttaccaggtactgtagttgacactaaaatattgtcatccaactgaattcgacttctacc
 tttgtagccatgcggtattcagggacaagcaggcctgcacattaccatgttctttggg
 acgagaacaatttcacagcagatggtattcaatctctgactaacaatctctgttatacct
 atgcgcggtgactcggtcggtctctatagtttctccagcgtattatgctcatcttgacg
 catttcgagcaggttctacactgggaacctgagataatgcaagacaacggatcaccgggtg
 aaaagaacacgaaaaacaactgtcgagacgtaggtgtgaagcctttaccagccttga
 aggagaatgtgaagagagtaattgttctactgctaaaaatccaaacattccttaatacagtt
 ttaataagtagtttggtgtgttctgttagttcggtcttagatttaccatgtttttctt
 atgtaaaatttgcggtttggttaagccttttaggaattagtgattagggtttttctaa
 agttgtacttttagctgatgataacggttgatgcagtgactttgttaaaacctcctctcta
 cagtagtgtttacgtcgttctc

>G1146 Amino Acid Sequence (domain in AA coordinates: 886-896)
 MPIRQMKDSSETHLVIKTQPLKHHNPKTVQNGKIPPPSPSPVTVTPATVTQSQASSPSP
 PSKNRSRRNRNGGRKSDQGDVCMRPSSRPKPPPPSQTSSAVSVATAGEIVAVNHQMOM
 GVRKNSNFAPRPGFGLTKCIVKANHFADLPKDLNQYDVTITPEVSSKSVNRATIAE
 LVRLYKESDLGRRLPAYDGRKSLYTAGELPFTWKEFSVKIVDEDDGIINGPKRERSYKVA
 IKFVARANMHLGEFLAGRADCPQEAQVILDIVLRELSVKRFCPVGRSFFSPDIKTPQR
 LGEGLBSWCGFYQSIRPTQMGSLNIDMASAAFIPLPVIEFVAQLLGKDVLSKPLSDSD
 RVKIKKGLRGVKEVTVHRANVRKRYRVALTTQPTRELMFPVDENCTMKSVIEYFQEMYG
 FTIQHTHLPCLQVGNQKKASYLPMACKIVEGQRYTKRLNEKQITALLKVTQRAEGQRN
 DILRTVQHNAYDQDPYAKEFGMNISEKLASVEARILPAPWLKYHENGKEKDCLPQVGQWN
 MMNKKMINGMTVSRWACVNFVRSVQENVARGFCNELGQMCEVSGMEFNPEPVIPIYSARP
 DQVEKALKHVYHTSMNKTGKLELELLAILPDNNGSLYGLDKRICETELGLISQCCLTKH
 VFKISKQYLADVSLKINVKMGRNTVLVDAISCRIPLVSDIPTIIFGADVTHPENGEES
 PSIAAVVASQDWPEVTKYAGLVCAQAHQELIQDLYKTWQDPVRGTVSGGMIRDLLISFR
 KATGQKPLRIIFYRDGVSEGFYQVLLYELDAIRKACASLEPNYQPPVTFIVVQKRHHTR
 LFANNHRDKNSTDRSGNILPGTVVDTKICHPTDFYLCSHAGIQGTSRPAHYHVLWDEN
 NFTADGIQSLTNNLCYTYARCTRSVSIVPPAYYAHLAAFRFRFYLEPEIMQDNGSPGKKK
 TKTTTVGVDVGKPLPALKENVKRVMFYC*

>G1267 (152..967)
 AAGTAGAGAATAATAATCACATCAAGATTGTTTATAACCTCCCCNTAATCACCTTCTTA
 NTNACCACCTCTCCGGCTCTCAACAGAACAAACAAAAACAGCTTCCGTTGTCTCTG
 TTCGGCGGAAATCGGACGGTCGAGATCAATCATGCATCGTAGAGCAGCAATCAAGAATC
 GGATGACGAAGAAGATGAGACTTACAACGACGTCGTTCTCTGAATCTCCTTCTTCTGTGA
 AGACTCAAAGATCTCAAAACCAACTCCAAAGAAAAGGAGGAACGTAGAGAAGAGAGTTGT
 CTCAGTTCCGATAGCTGACGTGGAAGGATCTAAGAGCAGAGGCGAAGTATATCCACCGTC
 CGATTCTATGGGCTGGAGAAAGTACGGACAAAAACCGATCAAAGGCTCGCCTTATCCCAG
 GGGATATTACAGATGTAGTAGCTCAAAGGATGTCCGGCGGAGGAAGCAGGTGGAGAGAAG
 CCGTGTGGACCTTCTAAGCTTATGATTACTTACGCCTGCGACCACAATCACCTTTCCC
 TTCTCTCTCCGCTAACACCAATCCCACCACCGCTCCTCCGTCGTCTCAAACCGCAAA
 GAAAGAGGAAGAATACGAAGAGGAGGAAGAAGAACTAACCGTCACCGCGCGAGAGGAACC
 ACCGGCGGACTTGATCTAAGCCACGTAGACTCACCGTTGCTATTAGGCGGCTGCTACAG
 CGAAATCGGAGAGTTCCGGTGGTTCTACGACGCGTCGATCTCATCATCTGGTTCTTC
 GAATTTCTCTCGACGTAACCTTAGAGAGAGGTTTTTCAGTAGGCCAAGAGGAAGATGAGTC
 TTTGTTCCGTGATCTCGGTGATTTACCTGATTGCGCCTCCGTGTTCCGCCGTGGGACTGT

TGCGACGGAGGAGCAACATCGAAGATGTGATTTTGGCGCCATTCTTTCTGTGATAGTTC
TAGATGAGTTTGTGTGTGTAGCCAAACCAAAGAAAAACACAATTTTTTTATTTTCC
ACTGTAAAGGTGTATCAATGGTGGATTCAATTTTTTAAAAA

>G1267 Amino Acid Sequence (domain in AA coordinates: 70-127)
MHRRAAIQESDDEEDETYNDVVPESPSSCEDSKISKPTPKKRRNVEKRVSVPIADVEGS
KSRGEVYPPSDSWAWRKYGQKPIKGSYPYRGYYRCSSSKGCPARKQVERSVDPSKLMIT
YACDHNHPFPSSSANTKSHRRSSVVLKTAKKEEYEEEEELTVTAEEPPAGLDLHVD
SPLLLGGCYSEIGEFWFYDASISSSSGSSNFDVTLERGFVSGQEEDESLFGDLGLDLPD
CASVFRRTGTATEEQHRRCDFGAIPFCDSSR*

>G1269 (88..951)

AACAATTCTCTCTCTTTATTCTTCTTCTTCAGCTTCAGATTTAGATCTTAAATCTTC
AAGTCTTCTTCTTCTTCTTCTGCAACCATGGCTATGCAGGAACGTTGTGAGAGTTTATGT
TCTGATGAACCTTATATCTTCTCAGATGCCTTTTACCTCAAGACAAGAAAGCCTTATACC
ATCACTAAACAAAGAGAGAAATGGACAGAAGCAGAGCATGAGAAGTTTGTAGAAGCATTG
AAACTCTATGGCAGAGCTTGGAGACGAATCGAAGAACATGTTGGAACAAAACTGCAGTT
CAGATTCGAAGCCATGCGCAGAAGTTCTTTACTAAGGTTGCTCGCGATTTTGGTGTAGC
TCTGAGTCCATTGAGATCCCGCTCCAAGGCCAAAGAGAAAGCCGATGCATCCTTACCCT
AGAAAGCTTGTGATTCCTGATGCAAAAGAGATGGTATACGCTGAACCTAACCGATCCAAG
CTGATTGAGGATGAAGATAACCGATCTCCAACATCGGTTTTATCAGCTCATGGCTCAGAT
GGATTAGGTTCCATTGGTTCAAATTCACCTAATCTTCTTTCAGCTGAGTTATCATCTCAC
ACAGAGGAATCATTTGCTCTCTAGAAGCAGAGACCAACAGAGCCCTTAAGCTCTTTGGAAA
ACTTTTGTAGTTGGTGATTACAATCTTCAATGAGTTGTGATGATTCTGAAGATGGCAAG
AAGAAGCTATACTCAGAAACACAGTCTTCTCAATGTTCTTCTTCTACTTCAGAAAACGCT
GAAACAGAAGTGGTAGTGTGCGAGTTCAAAGAAGTGAGAGATCAGCTTTCTCTCAGTTA
AAATCGTCGGTGACTGAGATGAACAACATGAGAGGGTTTCATGCCTTACAAAAGAGAGTA
AAGGTGGAAGAAAACATTGACAATGTAAATATCATATCCTTTGTGGTGAAGTGTTCGT
TTGTGTCAAGTCAGTTGTGTAACCTTTTGTGATCTCAACATCAGATTATGTGTATAATGT
CAGAGTATTAGGGAAAGTTTTTTTGGATTAGATTGTAAGATCACTCAAAGTTTCGTGT
CTTTCCATATAACCGATTAGAAATGAGATCCTTGTAACCTTAAACATTTTTATTGATCAA
TCAAATCTTCTGATGAAAAA

>G1269 Amino Acid Sequence (domain in AA coordinates: 27-83)
MAMQERCELSDELISSSDAFYLKTRKPYTITKQREKWTEAEHEKFVEALKLYGRAWRR
IEEHVGTKTAVQIRSHAQKFFTKVARDFGVSSSEIEIPPRPKRPMHPYPKRLVIPDAK
EMVYAEALTGSKLIQDENDRSPSTSVLSAHGSDGLSIGSNSPNSSSAELSSHTESLSLEA
ETKQSLKLFKTFVVDYNSMSCDSDSEDGKKLYSETQSLQCSSSTSENAETEVVSEF
KRSERSAFSQLKSSVTEMNMRGFMPYKKRVKVEENIDNVKLSYPLW*

>G1452 (175..1296)

ATTTATTAAGCATCAATGAGAGAACTTCAGAGCTGGGTTTGAGTTCTGTCCAATAATACA
TAACCACGTTATCATTTTTTGTCTTTACTATCTCATTACACTCTTCTGTTATTCGCCCAA
TTCTTACAGTCATTACTCTCTATAGGGCTCGAGCGGCCCGCCGGGAGGTTTCTATGCAG
ATGGTTACACTTCCCGCTCCATTGCCAGATTGGGTTTCGGTGTAAAGTCGCAATTAGTA
CTCACTATAGGGCTCGAGCGGCCCGCCGGGAGGTAAAGATCAAACAATGTCTAAAGAA
GCTGAGATGTGATCGCGGTGTCGGCTTTGTTCCCTGGTTTATGATTCTCTCTACTGAT
GTTGAACTTATCTCGTACTATCTTCGTCGTAATAATCGATGGTGTGAGAACTCTGTTGCT
GTGATTGCTGAGGTGAGATTACAAAGTTTCGAGCCGTGGGACTTGCCAGAGGAATCGAAA
CTGAAATCGGAGAACGAGTGGTTTTACTTCTGCGCGAGGGGAGGAAGTACCCGCACGGG
TCACAAAGCCGCGAGCCACACAGCTAGGATATTGAAAGCGACCGGTAAAGAGCGGAGT
GTTAAATCCGGAACCAAGTTGTTGGAACCAAGAGAACGCTTGTATTTATATCGGTCCG
GCTCCTCGTGGCGAGAGAACCGAGTGGATTATGCATGAATACTGCATCCATGGAGCCCCA
CAGGATGCATTAGTGGTGTGCCGGTTAAGAAAAATGCTGATTTTCGGGCTAGTTCGACC
CAAAAAATTGAGGATGGTGTGTGCAAGACGATGGCTACGTTGGCCAAAGAGGTGGTTTG
GACAAGGAGGACAAATCCTACTATGAATCTGAGCATCAGATACCAATGGTGACATCGCA
GAATCATCAATGTTGTTGAGGATCAGGCCGATACCGATGATGATTGTTACGCCGAGATT
CTGAACGATGATATAATAAGCTCGACGAAGAAGCGTTGAAAGCTAGCCAAGCGTTTCGA
CCAACCTAATCCAATCATCAAGAAACAATATCAAGCGAGTCATCGAGTAAGAGGTCAAAA
TGTGGTATAAAAAAAGAAATCAACGGAACAATGAATTGTTACGCTTTGTTTCAGGATCAAG
AACGTTGCCGAACCGACTCCAGCTGGAGATCCCGAACCCGTTCAAATCAAGAAAGAT

GATAGCCAGAGATTGATGAAGAATGTTCTGGCCACTACTGTTTTCTTGGCTATCTTATTT
TCTTTCTTTTGGACTGTATTAATAGCTAGGAACTAAAGCTAGTTACGACATACATATTAT
TTATACATAAATAAATATAGTATTTTGTCTATGGCAAAAAAAAAAAAAAAAAA

>G1452 Amino Acid Sequence (domain in AA coordinates: 30-177)
MQMVHTSRISIAQIGFGVKSQVLVTIGLERPPGQVKDQMSKEAEMSIASVAFPGFRFSP
TDVELISYYLRKIDGDENSVAVIAEVEIYKFEPWDLPEESKLKSENEWFYFCARGRKYP
HGSQSRRTQLGYWKATGKERSVKSGNQVVGTKRTLVPFHIGRAPRGERTEWIMHEYCIHG
APQDALVVCRLRNADFRASSTQKIEDGVVQDDGYVGQRGGLDKEDKSYESEHQIPNGD
IAESSNVVEDQADTDDDCYAEILNDDIIKLDEEALKASQAFRPTNPHTQETISSESSSKR
SKCGIKKESTETMNCYALFRIKNVAGTDSSWRFPNPFKIKKDDSQLMKNVLATTVFLAI
LFSFFWTVLIARN*

>G1494 (114..1406)

TCGACAGAGTTGTGTTGGGCGTGGAAC TTGGACTAGTTCCACATATCAGGTTATATAGAT
CTTCTCTTTCAACTTCTGATTTCGTCCAGAAGCTTTCCCTAATCTGAGATCTGACATGGAAC
ACCAAGGTTGGAGTTTGGAGGAGAATTATAGTTTGTCCACTAATAGAAGATCTATCAGGC
ACAAGATGAAGTATGAGTTATTATGGCGAGATGGACAAGTGGTTCTGCAGAGCCAAA
CTCATAGAGAACAAACCCAAACCAAGATCATCATGAAGAAGCCCTAAGATCCA
GCACCTTTCTTGAAGATCAAGAACTGTCTCTTGGATCCAATACCTCCAGATGAAGACC
CATTGCAACCCGACGACTTCTCTCCCACTTCTTCTCAACCATGGATCCCCTCCAGAGAC
CAACCTCAGAGACGGTTAAGCCTAAGTCCAGTCCCTGAACCTCCTCAAGTCATGGTTAAGC
CTAAGGCCTGTCTGACCTCCTCCTCAAGTCATGCCTCCTCCAAAATTTAGGTTAACAA
ATTCATCATCGGGGATTAGGGAACAGAAATGGAACAGTACTCGGTAACGACCGTTGGAC
CTAGCCATTGCGGAAGCAACCCATCACAGAACGATCTCGATGTCTCAATGAGTCATGATC
GAAGCAAAAACATAGAAGAAAAGCTTAATCCGAACGCAAGTTCTCATCAGGTGGCTCCT
CTGGTTGCAGCTTTGGCAAAGATATCAAAGAAATGGCTAGTGGAAAGATGCATCACAAACG
ACCGTAAGAGAAAACGTATAAATCACACTGACGAATCTGTATCTCTATCAGATGCAATCG
GTAACAAGTCGAACCAACGATCAGGATCAAACCGAAGGAGTCGAGCAGCTGAAGTTCATA
ATCTCTCCGAAAGGAGGAGGAGAGATAGGATCAATGAGAGAATGAAGGCTTTGCAAGAAC
TAATACCTCACTGCAGTAAACTGATAAAGCTTCGATTTTAGACGAAGCCATAGATTATT
TGAAATCACTTCAGTTACAGTTCAAGTGATGTGGATGGGAGTGGAATGGCGGCGGCGG
CGGCTTCGGCTCCGATGATGTTCCCGGAGTTCAACCTCAGCAGTTCATACGTCAGATAC
AGAGCCCGGTACAGTTACCTCGATTTCCGGTTATGGATCAGTCTGCAATTCAGAACAAATC
CCGGTTTAGTTTGGCAAACCCGGTACAAAACAGATCATCTCCGACCGGTTTGCTAGAT
ACATCGGTGGGTTCCCAACATGCAGGCCGCGACTCAGATGCAGCCGATGGAGATGTTGA
GATTTAGTTACCGGCGGGACAGCAAAGTCAACAACCGTCGTCTGTGCCGACGAAGACCA
CCGACGGTTCTCGTTTGACCCTAGGTTGGTGAGCCACTTTGC

>G1494 Amino Acid Sequence (domain in aa coordinates: 261-311)
MEHQGWSFEENYSLSNRRSIRPQDELVELLWRDGVVLSQTHREQTQTQKQDHHEAL
RSSTFLEDQETVSWIQYPPDEDPFEPDDFSSHFSTMDPLQRPSTETVKPKSSPEPPQVM
VKPKACPDPPPQVMPKPKFRLTNSSSGIRETEMEQYSVTTVGPSHCGSNPSQNDLDVSMS
HDRSKNIEEKLNPNASSSSGSSGCSFGKDIKEMASGRCTTDRKRKRINHTDESVSLS
AIGNKSNQRSGSNRRSRAAEVHNLSERRRRDRINERMKALQELI PHCSKTDKASILDEAI
DYLKSLQLQLQVMWMSGMAAAAASAPMMFPGVQPPQFIRQIQSPVQLPRFPVMDQSAIQ
NNPGLVCQNPVQNQIISDRFARYIGGFPHMQAATQMQPMEMLRFSSPAGQQSQQPSSVPT
KTTDGSRLDH*

>G1548 (1..2511)

ATGGCAATGTCTTGCAAGGATGGTAAGTTGGGATGTTTGGATAATGGGAAGTATGTGAGG
TATACACCTGAACAAGTTGAAGCACTTGAGAGGCTTTATCATGACTGTCTAAACCGAGT
TCTATTGCGCGTCAGCAGTTGATCAGAGAGTGTCTATTCTCTTAACATTGAGCCTAAA
CAGATCAAAGTGTGGTTTTCAGAACCGAAGATGTAGAGAGAAAACAAAGGAAAGAGGCTTCA
CGGCTTCAAAGCTGTGAATCGGAAGTTGACGGCAATGAACAAGCTCTTGATGGAGGAGAAT
GACAGGTTGCAGAAGCAAGTGTACAGCTGGTCCATGAAAACAGCTACTTCCGTCAACAT
ACTCCAAATCCTTCACTCCAGCTAAAGACACAAGCTGTGAATCGGTGGTGACGAGTGGT
CAGCACCATTGGCATCTCAAAATCCTCAGAGAGATGCTAGTCTGCAGGACTTTTGTCC
ATTGCAGAAGAACTTTAGCAGAGTTTCTTTCAAAGGCAACTGGAACCGCTGTTGAGTGG
GTTCAGATGCCTGGAATGAAGCCTGGTCCGATTCCATTGGAATCATCGCTATTTCTCAT
GTTGCACTGGTGTGGCAGCACGCGCCTGTGGCCTAGTGGGTCTTGAGCCTACAAGGGTT

GCAGAGATTGTCAAGGATCGTCCTTCGTGGTTCCGCGAATGTCGAGCTGTTGAAGTTATG
AACGTGTTGCCAACTGCCAATGGTGGAAACCGTTGAGCTGCTTTATATGCAGCTCTATGCA
CCAACCTACATTGGCCCCACACGCGATTCTTGGCTGTTACGTTACACCTCTGTTTTAGAA
GATGGCAGCCTTGTGGTGTGCGAGAGATCTCTTAAGAGCACTCAAAATGGTCCTAGTATG
CCACTGGTTCAGAATTTTGTGAGAGCAGAGATGCTTTCCAGTGGGTACTTGATACGGCCT
TGTGATGGTGGTGGCTCAATCATACACATAGTGGATCATATGGATTGAGGCTTGTAGC
GTGCCTGAGGTCTTGCGCCCGCTCTATGAGTCACCCAAAGTACTTGACAGAGAACAACA
ATGGCGGCACTGCGTCAGCTCAAGCAAATAGCTCAGGAGGTACTCAGACTAATAGTAGT
GTTAATGGGTGGGGACGGCGTCCTGCTGCCTTAAGAGCTCTCAGCCAGAGGCTAAGCAGA
GGCTTCAATGAAGCTGTAAATGGTTTCACTGATGAAGGATGGTCAGTGATAGGAGATAGC
ATGGATGATGTACAACTCACTGTAAACTCTTCTCCAGACAAGCTAATGGGTCTAAATCTT
ACATTTGCCAATGGCTTTGCTCCTGTAAGCAATGTTGTTTTATGCGCAAAAGCATCAATG
CTTTTACAGAAATGTTCTCCGGCGATCCTGCTTCGGTTTCTGAGGGAGCATAGGTGAGAA
TGGGCTGACAAACATTGATGCGTATCTAGCAGCAGCAGTTAAAGTAGGGCCTTGTAGT
GCCCCAGTTGGAGGATTGAGGGGAGGTTATACTTCCACTTGCTCATACTATTGAGCAT
GAAGAGTTTATGGAAGTCATCAAATTGGAAGGTCTTGGTCATTCCCCTGAAGATGCAATC
GTTCCAAGAGATATCTTCTTCTCAACTTTGTAGCGGAATGGATGAAAATGCTGTAGGA
ACCTGTGCGGAACCTATATTTGCTCCAATCGATGCTTCGTTTGGCGGATGATGCACCTCTG
CTTCTTCTGGTTTTCGTATTTATCCCTCTTGATTCCGCAAAGGAAGTATCTAGCCCAAAC
CGAACCTTGGATCTTGCTTCGGCACTGGAAATTGGTTCAGCTGGAACAAAAGCCTCAACT
GATCAATCAGGAACTCCACATGTGCAAGATCTGTGATGACAATAGCATTGAGTTTGGT
ATCGAGAGCCATATGCAAGAACATGTAGCATCCATGGCTAGGCAGTATGTTTCGAGGTATC
ATATCATCGGTGCAGAGAGTAGCATTGGCTCTTCTCCTTCTCATATCAGCTCAAGATT
GGTCTACGCACTCCTTTGGGTACTCCTGAAGCCCAAACACTTGCTCGTTGGATTGCCAG
AGTTACAGGGGCTACATGGGTGTTGAGCTACTTAAATCAAACAGTGACGGCAATGAATCT
ATTCTTAAGAATCTTTGGCATCACACTGATGCTATAATCTGCTGCTCAATGAAGGCCTTG
CCCCGTCTTACATTTGCAAACCAGGCGGGACTTGACATGCTGGAGACTACATTAGTTGCT
CTTCAAGACATCTCTTAGAGAAGATATTTGATGACAATGGAAGAAAGACTCTTTGCTCT
GAGTTCACACAGATCATGCAACAGGGCTTCGCGTGCCCTCAAGCGGGGATATGCTCTCA
AGCATGGGGAGACAGTTTCGTATGAGAGAGCAGTTGCTTGGAAAGTACTCAATGAAGAA
GAAATGCTCATTGCATCTGCTTTGTGTTTCATCAATTGGTCTTTGTGTGA
>G1548 Amino Acid Sequence (domain in AA coordinates: 17-77)
MAMSKDGKLGCLDNGKYVRYTPEQVEALERLYHDCPKPSSIRRQQLIRECPILSNIEPK
QIKVWFQNRRCREKQRKEASRLQAVNRKLTAMNKLMEENDRLQKQVSQLVHENSIFYROH
TPNPSPKADTSCSVVTSQGHLASQNPQRDASPAGLLSIAEETLAEFLSKATGTAVEW
VQMPGMPKPGPDSIGIIAISHGCTGVAARACGLVGLLEPTRVAEIVKDRPSWFRECRAVEVM
NVLPTANGGTVELLYMQLYAPTTLAPPRDFWLLRYTSVLEDGSLVVCERSLKSTQNGPSM
PLVQNFVRAEMLSSGYLIRPCDGGGSIHIVDHMDLEACSVPEVLRPLYESPKVLAQKTT
MAALRQLKQIAQEVTTQNSSVNGWGRPPAALRALSQLSRGFNEAVNGFTDEGWSVIGDS
MDDVTITVNSSPDKMLGLNLTFFANGFAPVSNVVLCAKASMLLQNVPPAILLRFLREHRSE
WADNNIDAYLAAAVKVGPC SARVGGFGQVILPLAHTIEHEEFMEVILEGLGHSPEDAI
VPRDIFLLQLCSGMDENAVGTCAELIFAPIDASFADDAPLLPSGFRIIPLDSAKEVSSPN
RTLDLASALEIGSAGTKASTDQSGNSTCARSVMTIAFEFGIESHMQEHVASMARQYVRGI
ISSVQRVALALSPSHISSQVGLRTPGLTPEAQTLARWICQSYRGYMGVELLKSNSDGNES
ILKNLWHHTDAIICCSMKALPVFTFANQAGLDMLETTTLVALQDISLEKIFDDNGRKTLC
EFPPQIMQQGFACLGIGLSSMRPVS YERAVAWKVLNEENAHICFPVFNWSFV*
>G1574 (1..1962)
ATGGATGATACAATGGACATGAGTTCAGGTAGTGATGAAGAAGTACAAGAAGAGAAGACC
ACTGTTAACGAGAGGGTCTATCATCAGGCTGCATTACAAGATCTGAAGCAACCCAAAGACC
GAAAAGGATCTACCTCCTGGTGTCTTACAGTTCTCTTATGAGGCATCAGAAAATTGCA
TTGAACTGGATGCGTAAGAAAAGAAAAAGAACAGGCACTGTTTGGGAGGGATATTAGCA
GATGATCAGGGACTTGGTAAACGATCTCGACGATCTCTTATCCTGTTACAAAAGTTG
AAGTCACAATCAAAGCAGAGAAAAGCGAAAAGGTCAAAACTCTGGTGGTACATTGATTGTT
TGTCCAGCAAGTGTGTAAACAATGGGCAAGAGAAGTTAAAGAGAAGGTTTCTGATGAA
CACAACTCTCTGTTTTAGTCCACCATGGATCTCACAGAACCAAGATCCACAGAAATA
GCAATATATGATGTGGTCATGACAACTTACGCCATTGTTACAAATGAAGTTCCACAAAAC
CCTATGCTGAATCGTTATGATAGTATGAGAGGCAGAGAAAGCCTTGACGGATCGAGTTTG

ATTACGCTCACGTTGGTGCCTAGGAAGAGTTAGGTGGTTGAGAGTAGTATTAGATGAA
GCTCATACAATTAACCAATAGAACCTTAATTGCAAAAGCTTGTTTTAGCCTTAGAGCC
AAAAGGAGATGGTGTGTTGACTGGAACGCCGATAAAGAACAAAGTAGACGATCTTTATAGC
TATTTAGATTTCTTAGATATCATCCATATGCCATGTGCAATTCATTTACCAAAGAATC
AAAGCTCCAATTGATAAAAGCCTCTTCATGGTTACAAGAAGCTTCAAGCTATTCTAAGG
GGTATAATGTTGCGCCGACCAAAGAATGGTCTTTCTACAGGAAGCTTGAATTGAATTCA
CGTTGGAAGTTTGAGGAATATGCTGCTGATGGGACTTTGTCATGAACACATGGCTTATCTT
TTGGTGATGCTTTTGCGACTACGCCAAGCTTGTAAACCATCCACAACCTGTTAACGGATAT
AGTCACTCAGATACTACAAGAAAAATGTCAGATGGAGTTCGAGTAGCCCCCTAGAGAGAAT
CTAATCATGTTCTCGATCTCTTGAAATTAATCCTCAACCACCTGCTCTGTTTGTAGTGAT
CCACCAAAGACCTGTTGTTACTTTGTGTGGCCATGTGTTTGTATGAGTGTGTGTCT
GTAAACATTAAACGGGGATAACAATACGTGCCCTGCACCTTAATTGCCACAGCCAGCTTAA
CATGATGTTGTTTTCACTGAATCTGCAGTTAGAAGTTGCATCAACGATTATGATGATCCT
GAAGATAAAAATGCTTTAGTTGCATCAAGCGAGTTTATTTTCATCGAAAAATCCGAGCTGT
GATAGAGATTCTTCAGTCGCTTGACAGCAAGGCAGTCCAGACACTCCACCAATAAAGAC
AATAGTATCAGTGGACTACAAGAAATCTCATTCTTACGTTTCTCAAAGACAAATGTAATGATTAT
GAAACAGGTGCGATGTTGATGTCTCTTAAAGCTGGAACCTTGGATTGAATATGGTAGCT
GCAAGTCATGTCATTCTACTGGACCTATGGTGAATCCAACAACAGAGGATCAAGCTATT
GATCGAGCTCATCGTATCGGACAACTCGAGCTGTTACGGTCACTCGTATTGCCATCAA
AATACCGTTGAGGAACGAATTTTGACTCTTCATGAACGTAAGGAACATTGTTGCATCT
GCATTGGGTGAAAAAACTGGCAAAAGTTCTGCGATTCAACTAACACTAGAAGATCTCGA
ATATCTGTTTTTTGGTGTGTAGAATATCCAGAGTTTTTATTGATAAGAGGAATAAAACC
TTAGCTATTTAATAAGTCACAAGTGTGAATGAATAA

>G1574 Amino Acid Sequence (domain in AA coordinates: 28-350)
MDDTMDMSSGSDEEVQEEKTVNERNVIYQALQDLKQPKTEKDLPPGVLTVPMLMRHQKIA
LNWMRKKEKRSRHLCLGGLIADDQGLKTIISTISLILLQKLKSQSKQRKRKGQNSGGTLIV
CPASVVQWAREVKEKVSDEHKLVLVHHGSHRTKDPTEIAIYDVVMTTYAIVTNEVPQN
PMLNRYDSMRGRESLDGSSLIQPHVGALGRVRLRVVLDEAHTIKNHRTLIAKACFSIRA
KRRWCLTGTPIKNKVDLYSYFRFLRYHPYAMCNSFHQRIKAPIDKPLHGYKKLQAILR
GIMLRRTKEWSFYRKLELNSRWKFEEYAADGTLHEHMAILLVMLRLRLQACNHPQLVNGY
SHSDTTRKMSDGVVRVAPRENLIIMFLDLLKLSSTTCSVCSDDPKDPVVTLCGHVFCYECVS
VNINGDNNTCPALNCHSQLKHDVVFTESAVRSCINDYDDPEDKNALVASRRVYFIENPSC
DRDSSVACRARQSRHSNTKDNSISGLNLIFTFLKDKCNDYETGAMLSLKAGNLGLNMVA
ASHVILLDLWNPPTTEDQAIIDRAHRIGQTRAVTVTRIAIKNTVEERILTLHERKRNIVAS
ALGEKNWQKFCDSNTNRRSRISVFWCVEYPRVFIIDKRNKTFSYLISHKCECNE*

>G1586 (1..807)
ATGAATCAAGAAGGTGCTTCACATAGCCCATCCTCCACTTCCACCGAACAGTCCGGGCA
CGTTGGTCACCTAAACCGGAGCAAATCTTGATACTCGAATCCATCTTCAACAGTGGTACT
GTTAACCACCAAAGATGAAACGGTGAGGATAAGAAAGATGCTTGAGAAATTCGGTGCT
GTGGGAGACGCAACGCTCTTCTACTGGTTTCAAACCGACGGTCAAGATCTCGCCGGAGA
CACCGGCAGCTTTTAGCAGCCACCACCGCAGCCGCCACCTCCATAGGAGCTGAAGACCAC
CAGCACATGACGGCCATGAGCATGCATCAATATCCTTGACAGCAACAACGAGATTGATTTG
GGTTTGGAAGTTGTAGCAACTTATCAGCTAATTACTTCTTAATGGATCGTCGTCATCT
CAAATCCCTTCTCTTTTCTCGGCCCTCTCTTCTTCAAGTGGTGGGTGTGAGAACAAAT
GGTATGGAGAATCTCTTCAAATGTATGGCCATGAATCTGATCATAATCATCAGCAGCAG
CATCATAGCTCAAATGCTGCATCAGTTTTAAACCCATCTGATCAAACTCCAACCTCCCAA
TACGAACAAGAAGGTTTATGACGGTGTATATAACGGAGTTCCTATGGAAGTAACAAAA
GGAGCAATAGACATGAAACAATGTTCCGGTGATGATTCGGTGTACTTCATTCCTCTGGT
CTTCTCTTCTCCACTGATGAGTTTGGTTCTTGTATGCATTCTTACAACATGGACAACT
TATTCTCTGGTACCGAGACAGACATGA

>G1586 Amino Acid Sequence (domain in AA coordinates: 21-81)
MNQEGASHSPSSSTSTEPVRARWSPKPEQILILESIENSGTVNPPKDETVRIRKMLEKFGA
VGDANVFYWFQNRSSRRRRHRQLLAATTAATSIGAEDHQHMTAMSMHQYPCSNNEIDL
GFGSCSNLSANYFLNGSSSSQIPSFLLGLSSSSGGCENNGMENLFKMYGHESDHNHQQQ
HHSSNAASVLNPSDQNSNSQYEQEGFMTVFINGVPMVETKGAIDMKTMFGDDSVLLHSSG
LPLPTDEFGFLMHSLOHQGTYFLVPRQT*

>G1786 (1..1170)

ATGATCGTGACGGTGGGGGAGCATCCGAGGACGGTGAAGGTGGAGGGGTGGTTCTGAAG
AAAGGGCCATGGACGGTGGCCGAGGACGAGACACTGGCGGCTTACGTACGGGAATACGGT
GAAGGGAACCTGGAATTCGTTCAGAAGAAGACATGGCTGGCTAGGTGTGGCAAGAGCTGC
CGCCTCCGCTGGGCTAACCCTTACGACCTAATCTCAGGAAAGGCTCCTTACCCCCGAG
GAAGAACGTCTCATACAACTCCACTCTCAGCTAGGCAACAAATGGGCTCGCATGGCT
GCTCAGTTACCAGGCAGAACAGATAACGAGATCAAGAACTACTGGAACACGAGGTTGAAA
CGCTTCCAACGCCAAGGCCTCCCTCTCTACCTCCAGAATATCCCAAACAATCATCAA
CAACAAATGTATCCTCAACAGCCCTCCTCACCTCTCCCGTCCCAAACACCTGCTTCTTCC
TTTACCTTTCTCTCTCCAACCGCCTTCTCTGTGTCCCAAACGTTGTTATAACACTGCC
TTCTCTCCCAAGGCCTCATATATTTCTTCTCCAACCAATTCCTTGTCTCGTCTCCGACC
TTCTTTCACACCCATTCTCTCTTCTCTCTATCAGTCTACCAATCCGGTTTACTCCATG
AAACATGAGCTCTCTTCAAACCAATTCCTACTCTGCCTCTTTAGGAGTCTATCAAGTA
AGCAAGTTCTCAGACAATGGGGATTGTAACCAAAACCTGAACACCGGTTTGCATACAAAT
ACCTGTCTCAGCTGTTAGAGGATCTTATGGAGGAGCCGAGGCTCTAGCTGATAGCTTTCGT
GCTCCTAAGCGGAGACAAATCATGGCTGCGCTTGAGGACAACAACAACAACAACACTTT
TTCTCGGGAGGTTTCGGACATCGTGTTCCTTCCAACAGTCTATGTTCTTGCAAGGTTTA
ACACCAAAGGAAGATGAGTCTCTCCAGATGAACACAATGCAAGATGAGGACATAACAAAG
CTTCTTGACTGGGGAAGTGAAGAAATCTCAAACGGGCAATCCTCTGTGATAACA
ACAGAGAACAACCTTGTCTTGACGATACCCAGTTCGCTTTTCTGTTTCCAGTTGATGAT
GACACCAACAACCTTGCCAGGGATCTGCTAG

>G1786 Amino Acid Sequence (domain in AA coordinates: TBD)

MIYVGGGASEDEGGGVVLKKGPTVAEDELAAAYVREYEGGNWNSVQKKTWLRCKGKSC
RLRWANHLRPNLRKGSFTPEBERLI IQLHSQLGNKWARMAAQLPGRDNEIKNYWNTRLK
RFQRQGLPLYPPEYSQNNHQQMYPPQPSPLPSQTPASSFTFLLQPPSLCPKRCYN TA
FSPKASYISSPTNFLVSSPTFLHTHSSLSYQSTNPVYSMKHELSSNQIPYSASLG VYQV
SKFSDNGDCNQNLNTGLHTNTCQLLEDLMEEAEALADSFAPKRRQIMAALEDNNNNNNF
FSGGFGHRVSSNSLCSLQGLTPKEDES LQNM TMQDEDITKLLDWGSESEIISNGQSSVIT
TENNLVLDDHQFAFLFPVDDDTNNLP GIC*

>G1792 (77..496)

AATCCATAGATCTCTTATTAATAACAGTGCTGACCAAGCTCTTACAAAGCAAACCAATC
TAGAACACCAAAGTTAATGGAGAGCTCAAACAGGAGCAGCAACAACCAATCACAAGATGA
CAAGCAAGCTCGTTTCCGGGGAGTTCGAAGAAGGCCTTGGGGAAAGTTTGCAGCAGAGAT
TCGAGACCCGTCGAGAAACGGTGCCCGTCTTTGGCTCGGGACATTTGAGACCGCTGAGGA
GGCAGCAAGGGCTTATGACCGAGCAGCCTTTAACCTTAGGGGTCATCTCGCTATACTCAA
CTTCCCTAATGATATTATCCACGTATGGACGACTACTCGCTTCGCCCTCCTTATGCTTC
TTCTTCTTCGTCGTCGTCATCGGGTTCAACTTCTACTAATGTGAGTCGACAAAACCAAAG
AGAAGTTTTTCAGTTTGAGTATTTGGACGATAAGGTTCTTGAAGAACTTCTTGATTGAGA
AGAAAGGAAGAGATAATCAGGATTAGTTTGTGTTTGATATTTATGTGGCACTGTTGTGG
CTACCTACGTGCATGATATGAGGTCGCTTGATTAGTACTTTATAACATGCATG
CCACGACCATAAATGTGAAGAGAAGACGTACTTTGCGTTTTCATGAAATATGAATGTTAG
ATGGTTTGAGTACAAAAAAAAAAAAAAAAAAAAA

>G1792 Amino Acid Sequence (domain in aa coordinates: 17-85)

MESSNRSSNNQSQDDKQARFRGVRRRPWGKFAAEIRDPSRNGARLWLGTFTAEAEARAY
DRAAFNLRLGHLAILNFPNEYYPMDYSLRPPYASSSSSSSSGSTSTNVSQRNQREVFEF
EYLLDDKVL EELL DSEERKR*

>G1865 (48..899)

AAGAAGAGGACATGAAGCACAGAGATTCTGCAGACTGCAGGTGACCAATGGACACTTTAT
CAATAAAAACATACTACTACTCTCTTACACTTTCAATTTTCCAATACAAATCCCAATCT
TTAATCTCTCTTTCTTCTCATCTCTCTTTCTCTTCTCTCTCATGGCTACAAGGATTC
CATTACAGAAATCACAATGGGAAGAACTTGAAACCAAGCTCTTGTGTTCAAGTACTTAG
CTGCAATATGCCTGTTCCACCTCATCTTCTTCTCATCAAAAGACCCCTTCTCTTCT
CTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCAAGCTTCTTCTCTCCACTCTTCTTCTCACT
TTGGGTGGAATGTGTATGAGATGGGAATGGGAAGAAAGATAGATGCAGAGCCAGGAAGAT
GTAGAAGAACTGATGGCAAGAAATGGAGATGCTCTAAAGAAGCTTACCCTGACTCTAAGT
ACTGTGAGAGACATATGCATAGAGGCAAGAACCCTTCTTCTCAAGAAAGCCTCCTCCTA
CTCAATTCACTCCAAATCTCTTTCTCGACTCTTCTTCCAGAAGAAGAAGTGGATACA
TGGATGATTTCTTCTCCATAGAACCTTCCGGGTCAATCAAAGCTGCTCTGGCTCAGCAA

TGGAAGATAATGATGATGGCTCATGTAGAGGCATCAACAACGAGGAGAAGCAGCCGGATC
GACATTGCTTCATCCTTGGTACTGACTTGAGGACACGTGAGAGGCCATTGATGTTAGAGG
AGAAGCTGAAACAAAGAGATCATGATAATGAAGAAGAGCAAGGAAGCAAGAGGTTTATA
GGTTTCTTGATGAATGGCCTTCTTCTAAATCTTCTGTTTCTACTTCACTCTTCATTGAT
CATCTTTTGTCTTATAACCTTGTATTTCTTGTTAAGATGGTAATGCAAATT
>G1865 Amino Acid Sequence (domain in AA coordinates: 124-149)
MDTSLIKTYLLSYTFNFIPIFIENLSFFFISLSLSLFMATRIPFTESQWEELENQALV
FKYLAANMPVPPHLLFLIKRPFLFSSSSSSSSSSSSFFSPTLSPHFGWNVYEMGMGRKIDA
EPGRCRRTDGKKWRCSKEAYPDSKYCERHMRGKNRSSSRKPPPTQFTPNLFLDSSRRR
RSGYMDDFFSIEPSGSIKSCSGSAMEDNDGSCRGINNEEKQPDHRHCFILGTDLRTREPR
LMLEEKLKQRDHDNEEQSGSKRFYRFLDEWPSSKSSVSTSLFI*
>G1886 (43..909)
AGGAAACATAAGTAATCGTTGCTTCGATCCTTTGTGTACATGGATGGATCCTGAACAGGAA
ATCTCAAACGAGACTTTGGAAACTATATTGGTAAGTTCAACAAAAGGAAGCAATAATAAC
AATAAGAAAATGGAAGAAGAAATGAAGAAGAAAGTATCAAGAGGAGAATTAGGAGGTGAA
GCTCAAAATTGTCCAAGATGTGAATCTCCAAACACAAAGTTTTGTACTACAACAACTAT
AGTCTCTCACAACCTCGTTACTTCTGCAAATCTTGTGCGGAGATATTGGACTAAAGCGCGT
ACTCTTCGTAACGTTCCCGTGGTGGTTGCCGTGCGAAACAAACGATCCTCTTCCTCA
GCTTTCTCCAAGAACAACAATAAGTCTATTAATTCCATACTGATCCACTTCAGAAC
CCTTTAATTACGGGAATGCCACCATCATCTTTTGGTTATGATCACTCCATTGATCTCAAC
CTCGCTTTCGCTACTCTCCAAAGCATCATTTATCTCTCAAGCTACTACGCCTTCTTTT
GGGTTTGGAGGTGATCTTTCTATTATGGAACCTCAACGAATGATGTAGGGATCTTCGGA
GGGCAAAACGGTACTTATAACAATAGTTTGTGTTATGGGTTTATGTCCGGAAATGGTAAT
AATAATCAAAATGAAATCAAGATGGCTTCTACATTGGGGATGTCTTGAAGGAAACGAG
AGAAAGCAAGAGAAATGTGAACAATAACAATAAATCAGAGAATCCTAGCAAGGTGTTT
TGGGGGTTTCCATGGCAGATGACCGGAGATTCCCGCGGAGTTGTACCGGAGATTGATCCC
GGAAGGGAAGCTGGAATGGGATGGTTTCATCTTGAATAATGGTTTACTCAACACTCCT
TTGGTCTAGCAGATCATTA
>G1886 Amino Acid Sequence (domain in aa coordinates: 17-59)
MDPEQEIENETLETILVSSTKGSNNNNKKMEEEMKKKVSRLGGEAQNCPRESPTNKF
CYNNYSLSQPRYFCKSCRRYWTGKGLLRNVPVGGGCRNRKRSSSAFKNMNNKSNINFH
TDPLQNPLITGMPSSFGYDHSIDLNLAFATLQKHLSSQATTPSFGFGDLISYGNSTN
DVGIFGGQNGTYNNSLCYGFMSGNGNNNQNNEIKMASTLGMSEGNERNKQENVNNNNNNSE
NPSKVFWGFPWQMTGDSAGVVEIDPGRESWNGMVSSWNNGLLNTPLV*
>G1933 (33..1418)
AATTGAGATTAAAGTAATTTATCTTTTCAGAAAATGGCGGTTGAAGACGATGTATCTTTGA
TAAGAACGACGACGTTAGTGGCACCAACAGACCCACGATTACAGTTCCATAGACCTC
CGCGCATCGAAACGGCGCGTATTTCTTTGGCGGTGGAGATGGGCTTAGTCTAAGCCCAG
GGCCACTTTCTTTTGTCTCTTTTGTGTTTGTGATAACTTCCCTGACGTCTTGACGCCGG
ATAACCAACGGACGACGTCGTTTACTCAGCTTCTTAACGGAATATGTCGGTGTCTCCTG
GTGGCGGAGGACGTTCAACGGCGGGGATGTTCCGCGGAGGAGGTCCGATGTTTACAATCC
CTTCTGGTTTCAGCCCTTCTAGTCTTCTCACCTCGCCCATGTTCTTTCCCCCGCAGTCGT
CAGCTCATACCGGCTTTATTCAACCACGGCAGCAGTCACAACCGCAACCACAACGACCAG
ACAGCTTTCCTCACCATATGCCACCATCGACATCCGTGCGCGTCCATGGTCGTCAATCTT
TAGACGTTTCACAAGTAGATCAAAGAGCTCGAAACCATATAATAATCCGGGGAATAACA
ATAATAACCGGTCGTATAACGTTGTGAACGTTGATAAACCGGCGGATGACGGTTATAACT
GGAGGAAGTACGGACAAAAGCCTATCAAAGGGTGTGAATATCCAAGGAGTTATTACAAAT
GTACACATGTTAACTGTCCGGTGAAGAAGAAAGTCGAACGGTCATCGGATGGACAGATCA
CTCAGATCATTTACAAAGGTCAACATGATCACGAGAGGCCTCAGAATCGCCGTGGCGGTG
GAGGCAGAGATTCCACTGAGGTTGGTGGTGCAGGGCAATGATGGAATCTAGTGATGATA
GTGGTTATCGTAAGGATCATGATGATGATGATGATGATGAAGATGATGAAGATCTTC
CGGCTTCAAAGATAAAGAATAAGACGGTGTGTGACGACTCACCGGACGGTGACCGGAGC
CTAAGATTATCGTTTCAGACAAAAGTGAAGTCGATCTTCTCGACGATGGCTATAGGTGGC
GTAAGTACGGACAAAAGTTGTCAAAGGAAATCCCCATCCAAGGAGCTATTATAAATGTA
CAACGCCAAATTGTACGGTCCGTAAACATGTAGAGAGAGCTTCCACGGATGCTAAGGCTG
TGATTACAACTTACGAAGGTAAACACAATCACGATGTCCCTGCCGCTAGAAACGGTACCG
CGGCAGCAACCGCAGCTGCGGTGGGGCCGTCTGACCACCATCGTATGAGATCAATGTCCG

>G1933 Amino Acid Sequence (conserved domain in AA coordinates:205-263, 344-404)

>G2059 (58..1089)

>G2059 Amino Acid Sequence (conserved domain in AA coordinates:184-254)

>G2105 (42..1487)

46

TTGGTGTTCATATGGAGAGGTTGATAGAGAAGCAAGATTTTTGGCTTGAGAAGTTGATGA
 AGATTGTGGAAGACAAAGAACATCAAAGGATGCTGAGAGAAGAGGAATGGAGAAGGATTG
 AAGCGGAAAGGATCGATAAGGAACGTTTCGTTTTGGACAAAAGAGAGGGAGAGGATTGAAG
 CTCGGGATGTTGCGGTGATTAATGCCTTGCAGTACTTGACGGGAAGGGCATTGATAAGGC
 CGGATTCCTCGTCTCTACAGAGAGGATTAATGGGAATGGAAGCGATAAAATGATGGCTG
 ATAATGAATTTGCTGATGAAGGAATAAGGGCAAGATGGATAAAAAACAAATGAATAAGA
 AAAGGAAGGAGAAATGGTCAAGCCACGGAGGGAATCATCCAAGAACCAAAGAGAATATGA
 TGATATACAACAATCAAGAACTAAGATTAATGATTTTTGTGCGAGATGATGACCAATGCC
 ATCATGAAGGTTACTCACCTTCAAACCTCAAGAACGCAGGAACCTCCGAGCTGCAGCAATG
 CCATGGCAGCTAGTACAAAGTGCTTTCCATTGCTTGAAGGAGAAGGAGATCAGAACTTGT
 GGGAGGGTTATGGTTTGAAGCAAAGGAAAGAAAATAATCATCAGTAAGCTACATTTTTCA
 TTCTCAAATGAAGAATAAGAGAACTTAGAAACGAT

>G2105 Amino Acid Sequence (domain in AA coordinates: 100-153)

MEDHQNHPPQYGIQPSQFSSDLFGFNLVSAPDQHHRLHFTDHEISLLPRGIQGLTVAGN
 NSNTITTIQSGGCVGGFSGFTDGGGTGRWPRQETLMLLEVRSLDHKFKKANQKGPLWDE
 VSRIMSEEHGYTRSGKKCREKFNLYKYKTKKKGKSGRRQDGKNYRFFRQLEAIYGESK
 DSVSCYNNTQFIMTNALHSNFRASNIHNIVPHHQNPLMTNTNTQSLSISNNFNSSSDL
 DLTSSEGNETTKREGMHWEKIKFVIGVHMERLIEKQDFWLEKLMKIVEDKEHQRLRE
 EEWRRIEAEERIDKERSFWTKERERIEARDVAVINALQYLTLGRALIRPDSSSPETERINGN
 SDKMMADNEFADEGNKGMKMDKKQMNKKRKEKWSHGGNHPRTKENMMIYNNQETKINDFC
 RDDDQCHHEGYSPPNSKNAGTPSCSNMAASTKCFPLLEGEQDNLWEGYGLKQORKENN
 H

Q*

>G2117 (49..465)

ATACTTGTCAACAAAAATTTCTTAAAGAACGCATAACTGTTTTTTCATGGCTGGTTCT
 GTCTATAACCTTCCAAGTCAAAACCCTAATCCACAGTCTTTATTCCAAATCTTGTGAT
 CGAGTACCACTTTCAAACCTTGCTGCCACGTCAGACGACTCTAGCCGGACTGCAGAAGAT
 AATGAGAGGAAGCGGAGAAGGAAGGTATCGAACCAGGAGTCAAGTCTCGGAGATCGCGTATG
 CGGAAACAGCGTCACATGGAAGAACTGTGGTCCATGCTTGTTCAACTCATCAATAAGAAC
 AAATCTCTAGTCGATGAGCTAAGCCAGCCAGGGAATGTTACGAGAAGGTTATAGAAGAG
 AACATGAAACTTCGAGAGGAAAACCTCAAGTTCGAGGAAGATGATTGGTGAGATCGGGCTT
 AATAGGTTTCTTAGCGTAGAGGCCGATCAGATCTGGACCTTCTAATCGTCTCGTAAGCTT
 GTTGGTTTTTGTGTTTATTTAAAG

>G2117 Amino Acid Sequence (conserved domain in AA coordinates: 46-106)

MAGSVYNLPQNPNPQSLFQIFVDRVPLSNLPATSDSSRTAEDNERKRRRKVSNRESAR
 RSRMRKQRHMEELWSMLVQLINKNKSILVDELSQARECYEKVIEENMKLRENSKSRKMIG
 EIGLNRFLSVEADQIWF*

>G2124 (87..923)

GAACAGCAAAACCCTAGATTTCTGTTCAAGCTCAAGACCGTACAAAACCTTGGAACCTCA
 TATATAAAGATCTCGAGAATAGCATTATGAATATCGTCTCTTGAAAGATGCAAACGACG
 AAGTTGCAGGCGGCGCTACGACAAGACGTGAAAGAGAAGTAAAAGAGGATCAAGAAGAAA
 CCGAAGTCAGAGCCACCAGTGGCAAAACCGTAATTAAAAAGCAGCCTACATCGATCTCTT
 CTTCTTCTTCTTCGTGGATGAAATCCAAGGATCCGAGGATTGTTAGGGTTTCACGCGCCT
 TTGGAGGCAAAAGACCGTCAAGCAAGTGTGTACGTTACGTGGACTACGTGACAGACGCG
 TGAGATTATCAGTCCCAACGCTATTTCAGCTCTACGATCTTCAAGAACGCTCGGTGTTG
 ACCAGCCTAGCAAAGCCGTTGACTGGTTGCTTGATGCAGCTAAAGAGGAGATCGACGAGC
 TACCTCCGTTACCTATCTCGCCGAAAATTCAGCATCTTCAACCATCATCAGTCCTTCT
 TGAATCTTGGTCAACCGCCCGTCAAGATCCGACCCAACTCGGGTTTAAATCAATGGAT
 GTGTACAAAAGTCTACTACTACTAGCCGCGAAGAAAACGATAGAGAGAAAGGAGAAAACG
 ATGTCGTTTACACAAACAATCATCATGTTGGGTCTTATGGAACCTATCACAACCTGGAAC
 ATCATCATCATCATCAACATTTGAGTTTACAGGCAGATTATCATAGTCATCAACTAC
 ATAGTCTTGTCCCATTTCCATCACAATTTTGGTATGTCCAATGACGACATCAACCAACAA
 CTACAACTATACAATCTTTGTTTCCATCATCATCGTCAGCTGGTTTCAGGGACTATGGAGA
 CATTAGATCCGAGGCAAATGTAGCAACAATGGTGGTAGAGACATTGATAATCGGATGTGCG
 TCGGTCCAATTCACCGAACTAATAGCACTACAACGGCTAACATGTCGAGGCATCTAGGC
 TCGGAGCGTTGTACAAGTAGAGGAAGTGATCACCATATGTGAAGTTAGATTATTGAAACG
 ATATAATTGTTGTTTGTATGTGTTTCAGAAATAAGGGGACAC

>G2124 Amino Acid Sequence (domain in AA coordinates: 75-132)

MNIVSWKDANDEVAGGATRREREVKEDQEETEVRATSGKTVIKQPTSISSSSSSSWMKS
KDPRIVRVSRAFGGKDRHSKVCTLRGLRDRRVRLSVPTAIQLYDLQERLGVDQPSKAVDW
LLDAAKEEIDELPPLPISPENFSIFNHHQSFLNLGQRPQDPTQLGFKINGCVQKSTTTS
REENDREKGENDVVYTNHHVGSYGTYHNLEHHHHHHQHLSLQADYHSHQLHSLVPPFSQ
ILVCPMTTSPTTTTIQSLFPPSSSSAGSGTMETLDPQM*

>G2140 (148..1254)

ACTCTCTTAACCTTTCTGTTTCTCTCTACCTTCTTTTACCAACCTTTCCTTTCTCTTACA
CACATATATATATACATATATAGAGAGAGAGAAGAGACAAAGAGTTGAAAGATGAAGAC
TCTCATGTCTTCATAGAAACAAGTGATATGTGCGCTAAGAAAGAAGAAGAAGAAGAA
GAAGAAGACAGTTCTGAAGCCATGAACAACATACAAAATTACCAAAATGACCTCTTCTTT
CACCAACTCATCTCTCATCATCACCATCATCATCATGATCCTTCTCAATCTGAAACTTTG
GGAGCATCCGGTAACGTTGGATCTGGTTTCACTATCTTCTCTCAAGATTCCGCTCTCTCCA
ATATGGTCTCTACCTCCACCTACCTCGATCCAACCACCATTTGATCAGTTTCTCTCTCT
TCTTCTTCTCCAGCATCTTCTACGGAAGTTTCTTCAACAGAAGTCGAGCTCATCATCAG
GGATTACAGTTTGGGTACGAGGGTTTGGTGGAGCCAGTCAGCAGCACATCATCATCAT
GAACAACCTCGGATCTTGTGCGAAGCTTTAGGTCCGGTAGTACAAGCCGGGTCCGGTCCT
TTTGGGTACAAGCTGAGTTAGGGAAGATGACAGCACAAGAGATCATGGACGCTAAAGCT
TTGGCTGCTTCAAAGAGTCATAGTGAAGCTGAGAGAAGAAGAAGAGAGAATCAATAAT
CATCTCGCTAAGCTCCGTAGCATATTACCCAACACCACCAAAACGGATAAAGCGTCGTTA
CTAGCTGAAGTGATCCAACATGTGAAAGAGTTGAAGAGAGAGACTTCAGTGATCTCAGAG
ACAAATCTTGTCCCAACGGAAGCGATGAGTTAACGGTAGCTTTCACGGAGGAGGAAGAA
ACCGGAGATGGCAGATTTGTAATTAAAGCGTCGCTTTGCTGTGAAGACAGGTCCGGATCTC
TTGCCGTGACATGATTAAACATTGAAAGCTATGCGTCTCAAACGCTCAAGGCGGAGATA
ACCACCGTTGGGGGACGAGTCAAGAACGTTTTGTTTGTACCGGAGAAGAGAGCTCCGGT
GAGGAAGTGGAGGAAGAGTACTGTATAGGGACGATTGAGGAAGCTTTGAAAGCGGTGATG
GAGAAGAGCAATGTAGAGGAATCATCTTCTTCTGAAATGCTAAGAGACAGAGAATGAGT
AGTCACAACACTATCACTATCGTCGAACAACAACAATATAATCAGAGGTAATCAATT
TTTTACTTAAATCGCTTTTTTTTCTTACTTTCGGTGTATCTACTACGTGTGTTGTTGCT
GGTTATGGAATGAATGTTGTACGTACGTTTACTATAGATATATGTGTGTTGTGTGT
ATGTATAACGGAAGTATTTGTATCCGTTGTGCTCTTGGACTTTTGGTTTGGTTCTAAGAT
ACTTATTTTTTAAAACTTGTATCGTTGAGTTGGTTTTCTAGATATGCTTAATGGGAGTAT
GTGACGAAAAAAA

>G2140 Amino Acid Sequence (domain in AA coordinates:167-242)
MCAKKEEEEEEDSSEAMNNIQNYQNDLFFHQLISHHHHHHHDPSQSETLGASGNVSG
FTIFSQDSVSPISLPPPTSIQPPFDQFPFSSSPASFYGSFFNRSRAHQGLQFGYEGF
GGATSAHHHHEQLRLISEALGPVVQAGSGPFLQAEKGMTAQEIMDAKALAASKSHSE
AERRRRERINNHLAKLRSILPNTTKDKASLLAEVIQHVKELKRETSVISETNLVPTESD
ELTVAPTEEEETGDGRFVIKASLCCEDRSDLLPDMIKTLKAMRLKTLKAEITTVGGRVKN
VLFVTGEESSEEEVEEYICIGTIEEALKAVMEKSNVEESSSSGNAKRQRMSSHNTITIVE
QQQQYNQR*

>G2144 (102..1241)

ATTAGGGTTTTGTTGTCGTGAGATTTGATTACACAAATTGCTGAATTTGGTTTCGATTAT
TGGTGTATTGTTTTTTCGAAGATTTCCAGTGAGTTTCCGTTTATGGATCTGACTGGAGGAT
TTGGAGCTAGATCCGGCGGTGTTGGACCGTGCCGGGAACCAATAGGCCTTGAATCGCTAC
ATCTCGGTGACGAATTTCCGGCAACTAGTGACGACTTTACCTCCCAGAACCCCGCGGTT
CGTTCACGGCTTTGCTTGAGCTTCCACCTACACAAGCAGTGGAGCTTCTCCATTTCACTG
ATTCTTCGTCTTCTCAACAAGCGGCAGTGACAGGGATCGGTGGAGAGATTCTCCGCCGC
TTCACTCTTTTCGGTGGGACATTGGCTTTTCTTCTAACTCAGTTCTCATGGAGCGAGCAG
CTCGTTTCTCGGTGATTGCCACTGAGCAACAAAACGGAATATCTCCGGGGAGACTCCGA
CGAGCTCTGTACCTTCCAATTCAGTGCTAATCTCGACAGAGTCAAGACGGAGCCTGCTG
AGACCGATTCTCATCTCAGCGTTGATTTCTGATTACGCGATTGAGAATCAAATCCCTTGCC
CTAACGAGAACAATCGAAATGGGAAGAGGAAAGATTTGAAAAGAAGGGTAAAGCTCGA
CGAAGAAGAACAAGGCTCTGAAGAGAACGAGAAGCTGCCATATGTTTACGTTAGAGCTC
GTCGTGGTCAAGCAACCGATAGCCATAGCTTAGCAGAACGAGCAAGAAGAGAGAAGATAA
ATGCACGAATGAAGCTGTTACAGGAACGGTCCAGGCTGTGATAAGATTCAAGGTACCG
CGCTGGTGTGAGATGAAATCAATACCATGTCCAGTCATTACAACGTCAAGTGGAGATGC
TATCAATGAGACTTGCTGCGGTAAACCCAGAAATCGACTTCAATCTCGACACCATATTGG

CTTCAGAAAACGGTTCTTTAATGGATGGGAGCTTCAATGCCGCACCAATGCAGCTTGCTT
GGCCTCAGCAAGCCATTGAGACCGAACAGTCCTTTTCATCACCAGCAACTGCAACAACCAC
CAACACAACAATGGCCCTTTTGACGGCTTGAACCAGCCGGTATGGGGAAGAGAAGAGGATC
AAGCTCATGGCAATGATAACAGCAATTTGATGGCAGTTTCTGAAAATGTAATGGTGGCTT
CTGCTAATTTGCACCCAAATCAGGTCAAATGGAGCTGTAAGTTGGGAAAACGGTAGAGA
TCATGAATGTGTATATACATCGTATAAGCTCGTTTCTCTCTATATAAATATAATCATAAA
TATAGATATCTGTTAAGAAGGTATCAGTCATTTGATTTCAGAGAGACAACACTGGTATGAT
TGTTTCTTATTCTGTACCAGATTTTCGACAATGTAGAATTTAGTAGGATATGATCATTTT
GATCTCGTTATATATA

>G2144 Amino Acid Sequence (domain in AA coordinates:203-283)
MDLTGGFGARSGGVGPCREPIGLES LHLGDEFRLVTTLPENPGGSFTALLELPPTQAV
ELLHFTDSSSSQAAVTGIGGEIPPLHSFGGTLAFPSNSVLMERAARFSVIAEQQNGN
ISGETPTSSVPSNSSANLDRVKTEPAETDSSQRLISDSAIENQIPCPNQNNRNGKRKDFE
KKGKSSTKKNKSSEENEKLPYVHVRARRGQATDSHSLAERARREKINARMKLLQELVPGC
DKIQGTALVLDEIINHVSQSLQRQVEMLSMRLAAVNPRIDFNLDTILASENGSLMDGSFNA
APMQLAWPQQAIETEQSFHHRQLQQPPTQQWPFQGLNQPVWGREEDQAHGNDNSNLMAVS
ENVMVASANLHPNQVKMEL*

>G2431 (47..1057)
CCCTTTCGTTTTTATTAAATTTCTTGGGTCGTTTCTTAAATTTGTATGTGTTTATTAAT
GGAGATCAACAATAATGCCAACATACTAATACTACTATTGATAATCACAAGGCAAAGAT
GAGCCTTGTTGTGTCAACGGATGCTAAGCCAAGGTGAAATGGACTTGTGATCTTCATCA
CAAATTCATCGAAGCCGTTAATCAACTTGGAGGACCTAACAAAGCAACACCTAAGGGTTT
GATGAAGGTTATGGAGATTCTTGGGCTTACCTTATACCATCTCAAGAGCCATTACAGAA
ATATCGGTTAGGGAAGAGCATGAAGTTCGATGATAACAAGCTAGAAGTTTCTCTGCATC
AGAGAATCAAGAAGTTGAGAGTAAAAACGATTCAAGAGATCTCCGAGGCTGCAGTGTCAC
CGAAGAAAACAGCAATCCAGCTAAAGAAGGGCTACAAATCACAGAGGCTTTACAAATGCA
GATGGAAGTTTCAGAAGAACTTCATGAACAAATCGAAGTTCAGAGGCATTTGCAGGTGAA
GATTGAGGCACAAGGAAAGTATCTACAGTCCGTTTAAATGAAAGCTCAACAACTCTCGC
TGGCTACTCATCTTCAAATCTCGGCATGGATTTTGGCAGGACCGAGCTCTCTAGATTAGC
TTCATGGTGGAACAGAGGCTGTCGAAGCACTTCGTTCTCAGAGCTAACGCAAGTAGAAGA
AGAAGAAGAAGGTTTCTTGTGTGTACAAGAAACAGAAACAGAGGAATTAGTCAGCTGAG
ATGTTTCAGTAGAGAGCTCGTTGACATCTTCAGAGACCTCAGAGACAAAACGGATACTGA
CAATAACCTTAATAAATCGATTGAACCTCCGTTGATGGAGATCAACTCGGAAGTGATGAA
GGGGAAGAAGAGAAGCATAAACGACGTCGTTTGGCTGGAGCAGCCTCTAATGAAGAGAGC
TTTTGGAGTTGATGATGATGAGCATTTGAAGTTGAGTTTGAATACTTACAAGAAAGACAT
GGAGGCGGTGTACGAACATAGGACTAGGGTTTAATTAATAAAAAAAACATTTTACTAAAGTT
ATATAAAATGTTTAAAGAATCCA

>G2431 Amino Acid Sequence (conserved domain in AA coordinates:38-88)
MCLLMEINNNANNTNTTIDNHKAKMSLVLSTDAKPRLKWTCDLHHKFIEAVNQLGPNKA
TPKGLMKVMEIPGLTLYHLKSHLQKYRLGKSMKFDDNKLEVSSASENQEVESKNDSRDLR
GCSVTEENSNPKEGLQITEALQMOMEVQKKLHEQIEVQRHLQVKIEAQGKYLQSVLMKA
QOTLAGYSSSNLGMDFARTELSRLASMVNRGCPSTSFSELTQVEEEEGFLWYKKPENRG
ISQLRCSVESSLTSSETSETKLDTDNNLNKSIELPLMEINSEVMKGKKRSINDVVCVEQP
LMKRAFGVDDDEHLKLSLNTYKDM EACTNIGLGFN*

>G2465 (86..1150)
CAATATTCTTTCTCCATTGAGATTAAGCTTCTTTCTCGCTGTCTCTCTATAGATCTT
GGTCTTAGTCCCTTTTGAATAATAATGATGGTGGAGATGGATTACGCTAAGAAAATGCA
GAAATGTGTCATGAATACGTTGAAGCACTTGAAGAAGAACAGAGAAAATCCAAGTCTTTCA
ACGCGAGCTTCCCTTTATGTTTAGAGCTTGTCACTCAAGCGATCGAAGCTTGTGCGGAAGGA
GTTATCTGGTACGACGACAACATACATCAGAACAGTGTTCAGAACAGACCACAAGTGTG
TGGTGGTCTGTCTTTGAAGAGTTTATTCTTATCAAGAAAATTAGTTCCTTGTGTGAAGA
AGTACAAGAAGAAGAAGAAGATGGTGAACATGAATCTTCTCCAGAACTTGTGAATAA
TAAGAAATCAGATTGGCTTAGATCTGTTTCAGCTATGGAATCATTACCGGATCTAAATCC
AAAAGAGGAGCGTGTAGCTAAGAAAGCGAAAGTGGTGGAGGTGAAACAAAAAGCGGTGC
GTTTCAGCCGTTTCAAAGCGCGTTTGGAGACTGATTTGCAACCGGCGGTGAAAGTAGC
TAGTTCGATGCCAGCGACGACGACGAGTTCACGACGGAACTTGTGGTGGTAAAGTGA
TTTGATTAAAGCTGGAGATGAGGAAAGACGGATAGAGCAGCAGCAATCGCAGTCGCATAC

GCATAGAAAACAAAGGCGGTGCTGGTCGCCGGAATTACACCGTCGATTCTTAAACGCGCT
 TCAGCAGCTTGGAGGATCTCATGTTGCTACACCAAAGCAAATCAGGGATCACATGAAGGT
 TGATGGATTAAACAAACGACGAAGTTAAAAGCCATTTACAGAAATATAGACTTCACACAAG
 AAGGCCAGCAGCAACATCCGTGGCGGCACAAAGTACCGGGAATCAGCAACAACCACAATT
 TGTGGTGGTTGGAGGCATATGGGTACCATCGTCACAAGATTTTCCACCACCGTCCGATGT
 AGCCAACAAGGTTGGTGTATATGCTCCGGTTGCGGTGGCGCAATCTCCAAAACGTTTCGTT
 GGAGAGAAGTTGCAACTCGCCGGCGGCATCTTCTCTACAAATACAAATACTTCTACTCC
 TGTGTATAAATCTGATAGTCACTATAATCATCTCTGATGTTGATTTTGGTGTAGGTT
 TGAATAATGTTTATGTGAATGTAA

>G2465 Amino Acid Sequence (conserved domain in AA coordinates:219-269)

MMVEMDYAKMKQKCHEYVEALEEEQKKIQVFQRELPLCLELVTQAIEACRKELSGTTTTT
 SEQCSEQTTSVCGGPVFEEFIPIKKISSLCEEVQEEEEEDGEHSSPELVNNKSDWLRS
 VQLWNHSPDLNPKERVAKKAVVEVKPKSGAFQFPQKRVLETDLPVAVKVASMPATTT
 SSTTETCGGKSDLIKAGDEERRIEQQSQSHTRKQRRCSPELHRRFLNALQQLGGS HV
 ATPKQIRDHMKVDVLTNDEVKSHLQKYLHTRRPAATSVAAQSTGNQQPQFVVVGGI WV
 PSSQDFPPPSDVANKGGVYAPVAVAQSPKRSLESCNSPAASSSTNTNTSTPVS*

>G2583 (38..607)

CAAATCAGAAAATATAGAGTTTGAAGGAAACTAAAAGATGGTACATTTCGAGGAAGTTCCG
 AGGTGTCCGCCAGCGACAATGGGGTTCTTGGGTCTCTGAGATTGCCCATCTCTATTGAA
 GAGAAGAGTGTGGCTTGAACCTTCGAAACGGCAGAAGCGGCTGCAAGAGCATACGACCA
 AGCGGCTCTTCTAATGAACGCCAAAACGCTAAGACCAATTTCCCTGTCTGTAATAATCAGA
 GGAAGGCTCCGATCACGTTAAAGATGTTAACTCTCCGTTGATGTCACCAAGTCATTATC
 TGAGCTTTTGAACGCTAAGCTAAGGAAGAGCTGCAAGACCTAACGCCTTCTTTGACGTG
 TCTCCGTCTTGATACTGACAGTTCCACATTTGGAGTTTGGCAGAAACGGGCCGGGTGCAA
 AACAAGTCCGACTTGGGTCATGCGCTCGAAGTTGGGAACGTAGTCAACGAAAGTGGCGT
 TGACTTAGGGTTGACTACGATGAACAAACAAAACGTTGAGAAAGAAGAAGAAGAAGA
 AGCTATTATTAGTGATGAGGATCAGTTAGCTATGGAGATGATCGAGGAGTTGCTGAATTG
 GAGTTGACTTTTGACTTTAACTTGTGCAAGTCCACAAGGGGTAAGGGTTTTTC

>G2583 Amino Acid Sequence (domain in AA coordinates:4-71)

MVHSRKFRGVRQRQWGSWVSEIRHPLLRVWLTGFETAEEAARAYDQAALLMNGQNAKT
 NFPVVKSEEGSDHVKDVNSPLMSPKLSSELLNAKLKSKDLTPSLTCLRLDTS SHIGV
 WQKRAGSKTSPTWVMRLELGNVNVESAVDLGLTMMNKQNVKEEEEEEEAIISDEDQLAME
 MIEELLNWS*

>G2724 (1..651)

ATGGAATAGAAATAAGGAGAGGTCCATGGACTGTGGAAGAAGACATGAAGCTCGTCAGT
 TACATTTCTCTTACGGTGAAGGATGGAACCTCCCTCTCTCGTTCTGCTGGACTGAAT
 AGAACGGGGAAAAGTTGCAGATTGCGGTGGCTAAATTATCTCCGGCCGGATATCCGCCGT
 GGAGACATATCCCTTCAAGAACAATTTATCATCCTTGAATCCATTCTCGTTGGGGAAAT
 CGGTGGTCAAAGATTGCTCAACATTTACCGGGAAGAACAGATAACGAGATAAAGAATTAT
 TGAGAACACGTTGTTCAAAAGCATGCAAAACTTCTAAATGTGACGTGAACAGCAAGCAA
 TTCAAAGACACCATCAAAATCTCTGGATGCCTCGTCTCATCGAGAGAATCGCCGCCACT
 CAAAGTGTCCAATTTACCTCTAACCCTACTCGCTGAGAACTCCAGCGTCGCCACCGCC
 ACGTCATCAACGTCGTCGTGAGGCTGTGAGATCGAGTTTCTACGGTGGTGATCAGGTG
 GAATTTGGAACGTTGGATCATATGACAAATGGTGGTTATTGGTTCAACGGCGGAGATACG
 TTTGAAACTTTGTGTAGTTTTGACGAGCTCAACAAGTGGCTCATACAGTAG

>G2724 Amino Acid Sequence (conserved domain in AA coordinates:7-113)

MEIEIRRGPTVEEDMKLVSYISLHGEGRWNSLSRSAGLNRTGKSCRLRWLNLYLRPDIRR
 GDISLQEQFIILELHSRWGNRWSKIAQHLPGRTDNEIKNYWRTRVQKHAKLLKCDVNSKQ
 FKDTIKHLWMPRLIBERIAATQSVQFTSNHYPENSSVATATSSSSSEAVRSSFYGGDQV
 EFGTLDHMTNGGYWNGGDTFETLCSFDELNKWLIQ*

>G377 (1..396)

atgggtctctcgcattttccaacagcgctcagaaggagtactaccacttctggtgatgaac
 acggttggtttcaatcactctgttgaagaacatggtgaggtctgttttcaaattggtgca
 tccgagactgaatctccatggagatagacgacgagcctgaagatgattttgtactaga
 agaactctcgataacacagttcaagtctctatgtgagaacatagaagaggaagaagaagag
 aaaggtgtggagtggtgtgtgtgcctttgtgggttaaagaggaagaggaagtgagtga
 ttggtttcttgcaagcatttctccacagagcttgtctagacaactggtttggttaataac

cacaccacatgccctctttgcaggtccattctctag

>G377 Amino Acid Sequence (domain in AA coordinates:85-128)
MGLSHFPTASEGVLPPLVMNTVVSITLLKNMVRSVFQIVASETESSMEIDDEPEDDFVTR
RISITQFKSLCENIEEEEEKGVCECCVCLCGFKEEEVSELVSKHFFHRACLDNWFGNN
HTTCPLCRSIL*

>G428 (97..1032)

TTACTTTTGTGTTTCTTCATATTCTTCAGAAGCAAGCACAAGGCTAGGGATCGAAGAAGC
GGCGATCACTGATCGTATCTCACTACGATCACATTAATGGATAGAATGTGTGGTTTCCGC
TCGACGGAAGACTATTCGGAGAAAGCGACGTTGATGATGCCGTCCGATTATCAGTCTTTG
ATTTGTTCAACCACCGGAGACAATCAAAGACTGTTTGGATCCGACGAACTCGCTACCGCT
TTGTCTCTCGGAGTTGCTTCCGCGTATTGAAAAGCTGAGGATAATTTCTCTCTTAGTGTC
ATCAAATCCAAATCGCTTCTCATCTTTGTATCCTCGCTTACTCCAAACCTACATCGAT
TGCCAAAAGGTGGGAGCGCCTATGGAAATAGCGTGTATATTGGAAGAGATTGAGCGAGAG
AACCATGTGTACAAGAGAGATGTTGCTCCATTATCTTGCTTTGGAGCTGATCCTGAGCTT
GATGAATTCATGGAAACCTACTGTGATATATTGGTTAAATACAAACCGATCTTGCGAGG
CCGTTTCGACGAGGCTACAACTTTTATAAACAAGATTGAAATGCAGCTTCAGAACTTGTGC
ACTGGTCCAGCGTCTGCTACAGCTCTTTCAGATGATGGTGGGTTTCATCTGACGAGGAA
CTGAGAGAAGATGATGACATAGCAGCGGATGACAGCCAACAAAGAAGCAATGACCGCGAT
CTGAAGGACCAGCTACTACGCAAATTTGGTAGCCATATCAGTTTCATTGAAACTCGAGTTC
TCTAAAAAGAAGAAGAAAGGAAGCTACCAAGAGAAGCAAGACAAGCGTTGCTCGATTGG
TGGAATGTTTCATAATAAATGGCCTTACCCTACTGAAGGCGACAAAATAGCTCTGGCTGAA
GAAACAGGTTTGGATCAAAAACAAATCAACAATTGGTTTATAAACCAAGGAAACGCCAT
TGGAAGCCTTCGGAGAACATGCCGTTTGATATGATGGACGATTCTAATGAAACATTCTTT
ACCGAGGAATGAAAAGAGAGACATGGGATTGTGCATTGTATAATTTTACACTGTTTTCC
CAAGAAAAGAAAACAGTAAAAGCTTTTGGTAAATGGGACATCATCGGAATGAATGGAA
CCAGTTAGCCAAAACGGTCAAGGGCGTGGCGTAACGAGACATTGTATTGGAAATAGTGGC
AATATTATGTCACTAATCTTCCAATGGTCCAAAATGATAGATTCTTATTTGTATTGAAC
CTTACTTAGATAGCTGATGTGTCAACTAAATAATTTATTTTCATCCTTATACTACTTGTA
TCAATGTCTCTAATTGATCAATTGTTGCTTGCTATTCAAAAAAAAAAAAAAAAAAAAAA

>G428 Amino Acid Sequence (domain in AA coordinates: 229-292)
MDRMCGFRSTEDYSEKATLMMPDSDYQSLICSTTGDNRQLFGSDELATALSSELLPRIRKA
EDNFSLSVIKSKIASHPLYPRLQLQTYIDCQKVGAPMEIACILEIQRENHVKRDVAPLS
CFGADPELDEFMETYCDILVKYKTDLARPFDEATTFINKIEMQLQNLCTGPASATALSDD
GAVSSDEELREDDIAADDSQQRSDNRDLKDQLLRKFGSHISSLKLEFSKKKKKGKLPRE
ARQALLDWNVHNKWPYPTEGDKIALAETGLDQKQINNWFINQRKRHWKPSENMPFDM
DDSNETFFTEE*

>G447 (241..3501)

CTTTTAAAGAGCTTAAAAAATTTGCTTTGAAGCTTCAAATATTCTTATGAACTAAAAAGAA
GAAAAAGCTTTTGTCTTTTCTTTCTTAGCAGCAGAATGATTTTGTCTTCAAAATATT
ACTATTTAGTTTCTCTCGTGCTCTTCTCTTGAGCAAATACAGATTCGTTAATTTTGCTGA
AGAAGAAGAACTCTGTTCTTCCCTGCACCAAACCAATTTTTCGTTCTTTCTATAAACC
ATGAAAGCTCCATCAAATGGATTTCTTCCAAGTTCCAACGAAGGAGAGAAGAAGCCAATC
AATTTCTCACTATGGCAGCGTTGTGCAGGGCCTTTAGTTTCATTACCTCCTGTGGGAAGT
CTTGTGGTTTACTTCCCTCAAGGACACAGCGAGCAAGTTGCAGCATCGATGCAGAAGCAA
ACAGATTTTATACCAAATTACCCAAATCTTCTTCTAAGCTGATTTGCTTGCTTCACAGT
GTTACATTACATGCTGATACCGAAACAGATGAAGTCTATGCACAAATGACTCTTCAACCT
GTGAATAAGTATGATAGAGAAGCATTGCTAGCTTCTGATATGGGCTTGAAGCTAAACAGA
CAACCTACTGAGTTTCTTGAAGACTCTTACTGCAAGTGACACAAGCACTCATGGTGGA
TTCTCTGTACCGCGTCGTGCAGCTGAGAAAATATCCCTCCTCTTGATTTCTCGATGCAA
CCGCTGCGCAAGAGATTGTAGCTAAAGATTTACATGATCTACATGGACTTTTCAACAT
ATCTATCGAGGCCAACCAAAAGACACTTGCTTACCACAGGTTGGAGCGTTTGTGTTAGC
ACAAAGAGACTATTTGCGGGTGATTCAGTTTGTGTTGTAAGAGATGAGAAATCACAGCTG
ATGTTGGGTATAAGACGTGCAATAGACAAACTCCGACTCTTCTCATCGGTATATCC
AGCGACAGTATGCACATTGGGATACTTGCAGCTGCAGCTCATGCTAATGCCAATAGTAGC
CCTTTTACCATCTTCTTCAATCCAAGGGCAAGTCCTTCAGAGTTGTAGTTCTTTAGCC
AAATACAACAAAGCCTTATACGCTCAAGTATCTTAGGAATGAGATTCGGGATGATGTTT
GAGACTGAGGATTGTGGGGTTTCGTAGATATATGGGTACAGTCACAGGTATTAGTGATCTT

GACCTGTAAAGATGGAAAGGCTCACAATGGCGTAATCTTCAGGTAGGATGGGATGAATCA
ACAGCTGGAGATAGGCCAAGCCGAGTATCCATATGGGAAATCGAACCCGTCATAACTCCT
TTTACATATGTCTCCTCCATTTTTCAGACCTAAGTACCCGAGGCAACCCGGGATGCCA
GATGATGAGTTAGACATGGAAAAATGCTTTCAAAAGAGCAATGCCTTGGATGGGAGAAGAC
TTTGGGATGAAGGACGCACAGAGTTCGATGTTCCCTGGTTTAAAGTCTAGTTCAATGGATG
AGTATGCAGCAAAACAATCCATTGTTCAGGTTCTGCTACTCCTCAGCTCCCGTCCGCGCTC
TCATCTTTTAACTTACCAACAATTTTGCTTCCAACGACCCCTCCAAGCTGTTGAACCTTC
CAATCCCCAAACCTCTCTTCCGCAAATTTCCCAATTCAACAAACCGAACACGGTTAACCAT
ATCAGCCAACAGATGCAAGCACAACCAGCCATGGTGAAATCTCAACAACAACAACAACA
CAACAACAACAACCAACACCAACAACAACAACCTGCAACAACAACAACAACCTACAGATG
TCACAGCAACAGGTGCAGCAACAAGGGATTTATAACAATGGTACGATTGCTGTTGCTAAC
CAAGTCTCTTGTCAAAGTCCAAACCAACCTACTGGATTCTCTCAGTCTCAGCTTCAGCAG
CAGTCAATGCTCCCTACTGGTGTCTAAAATGACACACCAGAACATAAATTTCTATGGGGAAT
AAAGGCTTGTCTCAAATGACATCGTTTGCGCAAGAAATGCAGTTTCAGCAGCAACTGGAA
ATGCATAACAGTAGCCAGTTTATTAAGAAACCGCAAGAACAGTCTCTCTCCATTCTTA
CAACAAAATCTGTCCCAAAATCCTCAGCAACTCCAAATGCAACAACAATCATCAAAACCA
AGTCTTTCACAACAGCTTCAGTTGCAGCTACTGCAGAAGCTACAGCAGCAGCAACAGCAG
CAGTCGATTCTCCAGTAAGCTCATCCTTACAGCCACAATTATCAGCGTTGCAGCAGACA
CAAAGCCATCAATTGCAACAACCTTCTGTGCTCTCAAAATCAACAGCCCTTGGCAGATGGT
AATAACAGCTTCCCAGCTTCACTTTTCATGCAGCCTCCACAGATTTCAGGTGAGTCTCAG
CAGCAAGGACAGATGAGTAACAAAAATCTTGTAGCCGCTGGAAGATCACATTTCTGGCCAC
ACAGATGGAGAGCTCCTTCTTGTTCAACTCACCTTCCGCAATAACACGGGACATGAT
AATGTTTTCACCGACAAATTTCTGAGCAGAAATCAACAGCAAGGACAAGCTGCATCTGTA
TCTGCATCTGATTAGTCTTTGAGCGCGCAAGCAATCCGGTCCAAGAGCTTTATACAAAA
ACTGAGAGCCGGATCAGTCAAGGCATGATGAATATGAAGAGTGTGGTGAACATTTCAGA
TTTAAAGCGCGGTAAACAGATCAAAATCGATGTATCCACAGCGGGAACGACGTACTGTCT
GATGTTGTTGGCCCTGTACAGCAGCAACAACTTTCCCACTACCATCATTGTTGTTGAT
GGAGATGCCAATCTCATCATCAAGAAACAACCTTAGCTTTCCCTGGTAATCTCGAAGCC
GTAACCTTCTGATCCTCTTCTCAAAAGGACTTTCAAACTTGGTTCCCACTATGGC
AACACACCAAGAGACATTGAGACGGAGCTGTCCAGTGCTGCAATCAGTTCTCAGTCATTT
GGTATTCCCAGCATTCCCTTTAAGCCCGGATGTTCAAATGAGGTTGGCGGCATCAATGAT
TCAGGAATCATGAATGGTGGAGGACTGTGGCCCAATCAGACTCAACGAATGCGAACATAT
ACAAAGGTTCAAAAACGAGGGTCAGTAGGTAGATCAATAGATGTTACCCGTTATAGCGGC
TATGATGAACCTTAGGCATGACTTAGCGAGAATGTTTGGCATCGAAGGACAGCTCGAAGAT
CCGCTAACCTCTGATTGGAACTCGTCTACACCGATCAGCAAAACGATATTTTACTAGTT
GGTGTATGATCCTTGGGAAGAGTTTGTGAATGCGTGCAACATAAAGATACTATCATCA
GTAGAAGTTAGCAAAATGAGCTTAGACGGAGATCTTGCAGCTATCCCAACCACAAACCA
GCCTGCAGCGAAACAGACAGCGGAAATGCTTGGAAAGTACACTATGAAGACACTTCTGCT
GCAGCTTCTTTCAACAGATAGAAATAAAAAGATGCAATATACCAAGTCAACTTACATTA
TCATTGAGGCCATCGCAAAGTACATGTTTCTTTTGTGTGTATGTACTGCAACAACAA
ACTGAGAAGAAGAAGATACTGCACGGTATATAAACATTTTATAGGACAGTGATTTGATT
TTTCATTCTAAGTTGATGTTGTTGACTTTCTTGTGTTCCATATTTGTATAACAAGTATAA
TGCTTGACAAGTCTATGAGGAGCATATCTTATACAGAGATACTAAGATGTAATGTTAATG
TAACTAAACAATTACCTTCATTAATCATGAATCCTTTGGTCTGTTTAAAA

>G447 Amino Acid Sequence (conserved domain in AA coordinates:22-356)

MKAPSNGLPSSNEGEKKPINSQWLWHACAGPLVSLPPVGSVLYVFPQGHSEQVAASMQKQ
TDFIPNYPNLPKSLICLLHSVTLHADTETDEVYAQMTLQPVNKYDREALLASDMGLKLN
QPTTEFFCKTLTASDSTHGGFSVPRRAAEKIFPPLDFSMQPPAQEIVAKDLHDTTWTFRH
IYRGQPKRHLITGWSVVFSTKRLFAGDSVLFVRDEKSQLMLGIRNRANRQTPTLSSSVIS
SDSMHIGILAAAAHANANSPTTIFFNPRASPSEFVVPLAKYNKALYAQVSLGMFRFRMF
ETEDCGVRRYMGTVTGISDLDPVRWKGSQWRNLQVGWDESTAGDRPSRVSIIWEIEPVITP
FYICPPPPFRPKYPRQPGMPDDELDMENAFKRAMPWMGEDFGMKDAQSSMFPGLSLVQWM
SMQQNNPLSGSATPQLPSALSSFNLPNNFASNDPSKLLNFQSPNLSSANSQFNKPNTVNH
ISQQMQAPAMVKSQQQQQQQQQHQHQQQQLQQQQQLQMSQQQVQQQGIYNNGTIAVAN
QVSCQSPNQPTGFSQSLQQQSMLPTGAKMTHQNINSMGNKGLSQMTSFAQEMQFQQQLE
MHNSSQLLRNQEQSSLHSLQQNLSQNPQQLQMQQQSSKPSPSQQLQLQLQLQKLQQQQQ
QSIPPVSSSLQPLSALQQTQSHQLQLLSSQNQQPLAHGNNSFPASTFMQPPQIQVSPQ

QQGQMSNKNLVAAGRSHSGHTDGEAPSCSTSPSANNTGHDNVSPTNFLSRNQQQGQAAASV
SASDSVFERASNPVQELYTKTESRISQGMNMKSAGEHFRFKSAVTDQIDVSTAGTTYCP
DVVGPVQQQQTFFPLPSFGFDGDCQSHHPRNNLAFPGNLEAVTSDPLYSQKDFQNLVPNYG
NTPRDIETELSSAAISSQSGFIPSIIPFKPGCSNEVGGINDSGIMNGGLWPNQTQRMRTY
TKVQKRGSVGRSIDVTRYSGYDELRLDLARMFGIEGQLEDPLTSDWKLVYTDHENDILLV
GDDPWEEFVNCVQNIKILSSVEVQMSLDGDLAAIPTTNQACSETDSGNNAWKVHYEDTSA
AASFNR*

>G464 (41..760)

CTCTGCTGGTATCATTGGAGTCTAGGGTTTGTATTGACATGCGTGGTGTGTCAGAATT
GGAGGTGGGGAAGAGTAATCTTCCGGCGGAGAGTGAGCTGGAATTGGGATTAGGGCTCAG
CCTCGGTGGTGGCGCGTGGAAAGAGCGTGGGAGGATTCTTACTGCTAAGGATTTTCCTTC
CGTTGGGTCTAAACGCTCTGTGTAATCTTCTCTCACCAGGAGCTTCTCTCTCGTTC
AAGTCAAGTGGTAGGATGGCCACCAATTGGGTTACACAGGATGAACAGTTGGTTAATAA
CCAAGCTATGAAGGCAGCAAGAGCGGAAGAAGGAGACGGGGAGAAGAAAGTTGTGAAGAA
TGATGAGCTCAAAGATGTGTCAATGAAGGTGAATCCGAAAGTTCAGGGCTTAGGGTTTGT
TAAGTGAATATGGATGGAGTTGGTATAGGCAGAAAAGTGGATATGAGAGCTCATTCTGTC
TTACGAAAACCTTGCTCAGACGCTTGAGGAAATGTTCTTTGGAATGACAGGTACTACTTG
TCGAGAAAAGGTTAAACCTTTAAGGCTTTTAGATGGATCATCAGACTTTGTACTCACTTA
TGAAGATAAGGAAGGGGATTGGATGCTTGTGGAGATGTTCCATGGAGAATGTTTATCAA
CTCGGTGAAAAGGCTTCGGATCATGGGAACCTCAGAAGCTAGTGGACTAGCTCCAAGACG
TCAAGAGCAGAAGGATAGACAAAGAAACAACCTGTTTAGCTTCCCTTCAAAGCTGGCA
TTGTTTATGTATTGTTGAGGTTTGCAATTTACTCGATACTTTTGAAGAAAGTATTTTG
GAGAATATGGATAAAAGCATGCAGAAGCTTAGATATGATTTGAATCCGGTTTTTCGGATAT
GGTTTTGCTTAGGTCATTCAATTCGTAGTTTTCCAGTTTTGTTTCTTCTTGGCTGTGTAC
CAATTATCTATGTTCTGTGAGAGAAAGCTCTT

>G464 Amino Acid Sequence (domain in AA coordinates: 20-28, 71-82, 126-142, 187-224)

MRGVSELEVGSNLPASELELGLGLSLGGGAWKERGRILTAKDFPSVSGSKRSAESSSHQ
GASPPRSSQVVGWPPIGLHRMNSLVNNQAMKAARABEGDGEKKVVKNDELKDVSMKVNPK
VQGLGFVKVNMDDGVGIGRKVDMRAHSSYENLAQTLLEEMFFGMTGTTCREKVKPLRLLDGS
SDFVLTYEDKEGDWMLVGDPWRMFINSVKRLRIMGTSEASGLAPRRQEQKDRQRNPNV*

>G557 (192..698)

CAGAGATCTGACGGCGGTAGCAGAGTAATCTATTCTCTCCCAAAATGTCTCGCAATTAGA
TTCTTTTCCAAGTTCTTCTGTAAATCCCAAGTCCCGCTCTTTTCTCTTTATCTTTTCAC
CAGCTTCGCTACTAAGACAACAAATCTTCCCTCTCTCTCTCGCCTGATCGATCTTCAA
GAGTAAGAAAAATGCAGGAACAAGCGACTAGCTCTTAGCTGCAAGCTCTTACCATCAA
GCAGCGAGAGGTATCAAGCTCTGCTCCACATTTGGAGATCAAAGAAGGAATTGAAAGCG
ATGAGGAGATACGGCGAGTGCCGGAGTTTGGAGGAGAAGCTGTGCGAAAAGAACTCCG
GTAGAGAATCTGGATCGGCGACCGGTGAGGAGCGGACACAGGCGACTGTGCGAGAAAGTC
AAAGGAAGCGAGGGAGGACACCGGCGGAGAAAGAGAACAAGCGGCTGAAGAGGTTGTTGA
GGAACAGAGTTTCAGCTCAGCAAGCAAGAGAGAGGAAAAAGGCTTACTTGAGCGAGTTGG
AAAAACAGAGTGAAAGACTTGGAGAACAAAACTCTGAACCTGAAGAGCGACTCTCTACTC
TTCAGAACGAGAACCAGATGCTTAGACATATTCTGAAGAACACAACAGGAAACAAGAGAG
GAGGTGGTGGTGGTTCTAATGCTGATGCAAGCCTTTGATCTCCTTCTTCTTGTGTTA
TATTTTTGTGATAAAATTTACAGAGAATTGTATCAATAATTATCATGTTAAAAATTATAT
GGGATGTGAGAGCTAATATTGCAATTGTAGACCAAGTTCTCTTAAAAAATAAAAAAA
AA

>G557 Amino Acid Sequence (domain in AA coordinates: 90-150)
MQEQATSSLAASSLPSSSSSSAPHLEIKEGIESDEEIRRVPFEGGEAVGKETSGRES
GSATQOERTQATVGESQQRKGRTPAEKENKRLKRLLRNVSAQQARERKKAYLSELENRV
KDLENKNSELEERLSTLQENQMLRHILKNTTGNKRGGGGSNADASL*

>G577 (44..2155)

AAAAACAGACTGAGAGAGAGAGAGAGAGTGTGTTGTTGGCCATGGGATGCACGGCCTC
CAAGCTCGACAGTGAGGATGCTGCTCCGTCGCTGCAAGGAGCGCGCGCTTATGAAGGA
CGCGTCTACGCTCGTCACCATCTCGCCGCGCTCACTCTGACTACTGCCGCTCCCTTCG
TCTCACTGGCTCTGCCCTCTCTCTCTCGCCGCGGAGCCCTCTCCGCTCTCCGAGAA
TACTCCCGCTGTTTTCTCCGCCCTTCTCCAGTCAGGACGCGCCACGTGTCCCTTCTTC

CCATTCCCCAGAACCCCCCTCCTCCGCCCATCCGCAGCAAGCCTAAGCCTACTAGGCCTAG
 GAGGCTTCCACACATTCTCTCCGACTCCTCTCCTTCTTCTCTCCTGCCACCAGTTTCTA
 TCCCACCTGCTCACCAGAACTCTACTTACTCTCGCTCTCCATCTCAAGCTTCTCTGTCTG
 GAACTGGGAGAATTTCTACCTCTCCTCTCCCCCGACTCCGAGTACTTCGAACGCAAAGC
 TCGCCAGAACCACAAGCACCGTCTCTTCCGACTACGACGCCGAAACTGAAAGATCCGA
 CCACGATTACTGCCACTCACGGAGAGATGCCGCCGAGGAAGTTCACTGCAGCGAGTGGGG
 CGACGACCACGACCGTTTCACTGCCCACCTCTTCTCGTCCGACGGAGATGGGGAGGTGAAAC
 TCACGTTTCCAGATCCGGTATTGAAGAAGAGCCTGTGAACAACACATCAAGACCCAAA
 TGGCAAAGAGCACTCTGACCATGTTACCACCTTCTTCCGACTGCTACAAGACCAAATTGGT
 GGTAAAGGCACAAGAATTTGAAGGAGATCCTTGACGCCGTTCAAGACTACTTCGACAAGGC
 TGCCTCCGCTGGGGACCAGGTCTCCGCCATGCTTGAGATCGGCCGGGTGAGCTCGACCG
 CAGCTTCAGCAAGCTGAGGAAGACGGTGTATCATTCAAGCAGTGTGTTAGCAACTTGAG
 CGCAAGCTGGACCTCAAAACCCCCATTGGCAGTCAAATACAAGCTCGATGCATCTACCTT
 GAATGATGAACAAGGCCGGCCTCAAGAGCCTTGTCTCACTCTAGACCGACTCCTCGCTTG
 GGAAGAAGCTTTATGAGGATGTCAAGGCAAGAGAAGGAGTTAAGATTGAGCACGAGAA
 GAAGCTGTCTGCGCTGCAGAGTCAGGAGTATAAGGGAGGTGATGAATCCAAGCTAGACAA
 GACTAAACTTCCATAACAGACTGCAATCACTCATCATTGTTTCTTCAGAAGCTGTTTT
 AACCACGTCTAATGCCATTCTCCGCCCTCCGGGACACTGACCTTGTCCCTCAGCTTGTGA
 ACTCTGCCACGGATTAATGTACATGTGGAAGTCAATGCACGAGTATCAGGAAATCCAGAA
 CAACATCGTGCAACAAGTCCGTGGCCTGATCAACCAACAGAGAGAGGTGAGTCAACATC
 AGAGGTACACCGGACGGTGACCGGGACCTAGAGTCAGCTGTGTCTTGTGGCATTTCGAG
 CTTCTGTGCGCATCATTAATTCAGAGGGAGTTCATATGCTCTCTCCACGATGGTTCAA
 GCTGAGCCTGGTTCCCTGAGCAACGGAGACCCAAAGAAACAGCGGCCAGACTCATTTCG
 CTTGTGCGAGGAGTGGAAGCAGAGCCTGGAACGGGTGCCTGACACAGTGGCGTCAGAAGC
 CATAAGAGCTTTGTAAACGTGGTACATGTGATATCAATAAGCAGGCGGAAGAGGTGAA
 GATGAAGAAACGCACGGAGAGTGCAGGAAAGGAGCTGGAGAAGAAAGCATCCTCAGTGA
 GAGCATAGAGAGGAAGTACTACAGGCATACTCGACGGTGGGATAGGCCCTGGACCCGA
 GGTGTTGGACTCACGGGACCCGCTATCTGAGAAGAAATGTGAGCTGGCGGCATGTCAGAG
 GCAGGTGGAGGATGAGGTAATGAGGCACGTGAAGGCTGTGGAGGTGACACGAGCTATGAC
 TCTCAACAATCTACAAACCGGCCTGCCAATGTATTCCAGGCCTTGACCAGCTTCTCATC
 TCTCTTCACTGAATCTCTCCAGACTGTCTGTTCTCGTTCCTACTCCATCAACTGATTATG
 TCCAAGTTTCTCATTTATTTTAAAGCTCTCATTACGTGTGATCATGTAATTTGAGGAT
 TGATTAAATTGAGTCTTGTGGTTTTGTGAGGACTCACAATCTTCTCATTAAAAA
 AAAAAAAAAA

>G577 Amino Acid Sequence (domain in AA coordinates: TBD)
 MGCTASKLDSSEDAVRRCKERRRLMKDAVYARHHLAAHSDYCRSLRLTGSALSSFAAGEP
 LSVSENTPAVFLRPSSSQDAPRVPSSSHSPEPPPPPIRSKPKPTRPRRLPHILSDSSPSSS
 PATSFYPTAHQNSTYSRSPSQASSVWNWENFYPPSPDSEYFERKARQNHKRPSPDYDA
 ETERSDDHYCHSRDDAAEEVHCSEWGDDHDFRFTATSSSDGDGEVETHVSRSGIEEFPVKQ
 PHQDPNGKEHSDHVTSSDCYKTKLVVRHKNLKEILDVQDYFDKAASAGDQVSAMLEIG
 RAELDRSFSKLRKTVYHSSSVFNSLSASWTSKPPLAVKYKLDASTLNDEQGLKSLCSTL
 DRLLAWKKLYEDVKAREGVKIEHEKKLSALQSQYKGGDESKLDKTKTSITRLQSLIIV
 SSEAVLTTSNAILRLRDTDLVPQLVELCHGLMYMWSMHEYHEIQNNIVQQVRGLINQTE
 RGESTSEVHRQVTRDLESASVSLWHSSFCRIIKFQREFICSLHAWFKLSLVPLSNGDPKKQ
 RPDSPALCEEWKQSLERVPDTPVASEAIKSFVNVVHVVISIKQAEVVKMKRTESAGKELEK
 KASSLRSIERKYQAYSTVGIGPGEVLDSDRDLSEKKCELAACQORQVEDEVMRHVKA
 VTRAMTLNQLTGLPNVQALTSFSSSLFTESLQTVCSRYSIN*

>G674 (1..786)-

ATGGTGTTTAAATCAGAAAAATCAAACCGGGAATGAAATCAAAGGAGAAGCAAAGGAAG
 GGATTATGGTCACCCGAGGAAGATGAGAAGCTTAGGAGTCATGTCTCAAATATGGCCAT
 GGATGCTGGAGTACTATTCTCTCAAGCTGGATTGCAGAGGAATGGGAAGAGTTGTAGA
 TTAAGGTGGGTTAATTATTTAAGACCTGGACTTAAGAAGTCTTTATTCACTAAACAAGAG
 GAACTATACTTCTTTCACTTCATTCCATGTTGGGTAACAAATGGTCTCAGATATCGAAA
 TTCTTACCAGGAAGAACCGACAACGAGATCAAAACTATTGGCATTCTAATCTAAAGAAG
 GGTGTAACCTTTGAAACAACATGAAACCAAAAAACATCAACACCTTTAATCACAAC
 TCACTTGAGGCCCTGCAGAGTTCAACTGAAAGATCTTCTCATCTATCAATGTGCGAGAA
 ACGTCTAATGCTCAACCTCAAGCTTTTCGCCAATCTCGTGTCTCGGAATGGTTAGAT

CATAGTTTGCTTATGGATCAGTCACCTCAAAAGTCTAGCTATGTTCAAAATCTTGTTTTC
CCGGAAGAGAGAGGATTATTGGACCATGTGGCCCTCGTTATTTGGGAAACGACTCTTTG
CCTGATTTTCGTGCCAAATTCAGAAATTTTGTGGATGATGAGATATCATCTGAGATCGAG
TTCTGTACTTCATTTTCAGACAACTTTTTGTTCGATGGTCTCATCAACGAGCTACGACCA
ATGTAA

>G674 Amino Acid Sequence (domain in AA coordinates: 20-120)
MVFKEKSNREMKSKKEQRKGLWSPEEDEKLRSVHLKYGHGCWSTIPLQAGLQRNGKSCR
LRWVNYLRPGLKKSLFTKQETILLSLHSLGNKWSQISKFLPGRDNEIKNYWHSNLKK
GVTLKQHETTKKHQTPLITNSLEALQSSTERSSSSINVGETSNAQTSSFSPNLVFSEWLD
HSLMDQSPQKSSYVQNLVLPPEERGFIPGCPRYLGNDSLPDFVPNSEFLDDEISSEIE
FCTSFSDNFLFDGLINELRPM*

>G736 (1..513)
ATGGCGACTCAAGATTCTCAAGGGATTAACTCTTTGGCAAACTATTGCATTTAACACT
CGAACAAATAAAAAATGAAGAAGAGACACACCCGCCGGAGCAAGAAGCCACAATAGCCGTT
AGATCATCATCATCATCGGATCTGACGGCCGAGAAGCGTCCGGATAAGATCATAGCATGT
CCAAGATGCAAGAGCATGGAGACAAAGTTCTGTTACTTCAACAACACGGAATCAG
CCTCGACACTTTTGTAAAGGCTGCCACCGTTACTGGACCGCCGGTGGTGCCTCCGGAAC
GTTCCCGTCCGGCCCGGTCGTCGGAAGTCCAAACCACCTGGTCGTGTCGTTGGTATG
CTTGGAGATGGAAATGGTGTTCGCCAAGTCGAGCTTATAAATGGCTTGCTCGTTGAGGAG
TGGCAGCATGCCGACCGCAGCTCACGGTAGTTTCCGGCATGATTTTCCCATGAAGCGG
TCCGGTGTACTCCGACGGTCAATCGTGTGA

>G736 Amino Acid Sequence (domain in AA coordinates: 54-111)
MATQDSQGIKLFKGTIAFNTRTIKNEETHPPEQEATIAVRSSSSDLTAEKRPDKIIAC
PRCKSMETKFCYFNNGNQRHFCCKGCHRYWTAGGALRNVVPGAGRRKSKPPGRVVVGM
LGDGNGVRQVELINGLLVEEQHAAAAAHGSFRHDFPMKRLRCYSDGQSC*

>G903 (96..1496)
CCCCGGTCCGACCGCGTCTCTCTCTGAACTATACAAAAACCTACTTTTAAT
TTCTCTTCCAAGAGTCAAGAACCCAGAAAGACATGACAAGTGAAGTTCTTCAAACAA
TCTCAAGTGGATCAGGTTTTGCTCAGCCACAGAGCTCATCAACCCTGGATCATGATGAAT
CTCTCATCAATCCTCTCTTGTAAAGAAAAGAGAAATCTCCCTGGAAATCTGATCCGG
AAGCTGAAGTGATAGCTTTATCCCCACGACCTTGATGGCTACGAACCGGTTCTATGTG
AGGTATGTGGCAAAGGTTTCCAAAGAGACCAAACTTACAGCTTCATCGGCGAGGACATA
ATCTTCCATGGAAGTTGAAGCAGAGGACAAGCAAAGAAGTGAGAAAACGTGTCTACGTTT
GCCCCGAGAAGACATGTGTCCACCATCACTCCTCTAGAGCTCTAGGCGATCTCACTGGAA
TCAAAAAGCATTTTTGCCCCGAAACACGGGGAGAAGAAGTGAGCTGCGAGAAATGTGCTA
AGAGATACGCAGTCCAATCTGATTGGAAGCTCATTTCAAGACTTGTGGTACTAGAGAGT
ACCGTTGCGATTGTGGCACCATTCTCAAGGCGAGACAGCTTTATCACTCATAGAGCTT
TCTGCGATGCCTTAGCGGAAGAAACCGCTAAGATAAACGCAGTGTCTCATCTCAACGGTT
TAGCCCGCGGCTGGAGCCCCAGGATCAGTTAATCTCAACTATCAATATCTCATGGGAACAT
TCATCCCACCGCTTCAACCATTTGTACCACAACCGCAAACAAATCCAAACCATCATCATC
AACATTTTCAGCCACCAACTTCTTCGTCGCTCTCTCTATGGATGGGACAAGATATCGCGC
CGCTCAACCGCAACCGGACTACGATTGGGTTTTTGGAAACGCTAAGGCAGCGTCTGCTT
GCATTGATAATAATAACTACGATGAGCAGATTACGCAAAACGCAACGCAAGTTTGA
CCACTACCACTACTCTCTGCCCCCTCTTTATTCAGCAGCGACCAACCACAAAACGCAA
ACGCAAAATCAACCGTGAATATGTCCGCGACAGCTTTACTACAGAAAGCTGCTGAAATG
GCGCTACTTCTACAACAACCGCAGCGACCAATGACCCATCAACGTTTCTTCAAAGTTTCC
CGCTTAAATCCACCGATCAAACCACCAAGTTATGACAGTGGCGAAAAGTTTTTGTCTTGT
TCGGGTCTAACAACAACATTGGGTTAATGAGTCGTAGTCATGATCATCAAGAGATCGAGA
ACGCTAGAAATGACGTTACGGTTGCGTCTGCTTGGATGAATTACAGAATTACCCCTTGA
AACGTAGAAGAGTTGATGGTGGAGGTGAAGTGGGTGGAGGAGGGCAAACTCGGGATTTC
TCGGGGTTGGGTGTACAAACGTTGTGCCATCCATCGTCTATCAATGGATGGATTGAAAGA
GTTTAAAAATTTTCGGGTTAATGCATAAATTACGTAAAAGAAGAAGGAATCTTTTTCAT
TTCCACCATTTTCTAAGATAACATATGTATATGGTAATGGAAGTTGTTTTCTTTTATTAA
TTCAATATTCTAAACTTATGATATATGTATAATGAATGTGTTTATCTTCAA

>G903 Amino Acid Sequence (domain in AA coordinates: 68-92)
MTSEVLQTISSSGSGFAPQSSSTLDHDESLINPPLVKKRNLPGNPDPEAEVIALSPPTL
MATNRFLCEVCGKGFQRDQNLQLHRRGHNLPWKLKQRTSKEVRKRVYVCPEKTCVHHSS

RALGDLTGIIKKHFCRKHGEKKWTCEKCAKRYAVQSDWKAHSKTCGTREYRCDGTFISRR
 DSFTTHRAFCDALAEETAKINAVSHLNGLAAGAPGSVNLNYQYLMGTFIPLQPFVPPQ
 QTNPNHHQHFPPTSSSLSLWMGQDIAPPQPQPDYDWVFGNAKASACIDNNTHDEQI
 TQNNANASLTNTTSLAPSLFSSDQPNANANSNVNMSATALLQKAAEIGATSTTTAATND
 PSTFLQSFPLKSTDQTTSYDSGEKFFALFGSNNNIGLMSRSHDHQEIENARNVDVTASAL
 DELQNPWKRRRVDGGGEVGGGGQTRDFLGVGVQTLCHPSSINGWI*

>G917 (32..679)

TTAGGGTTTTAGAAAGATAGATCGATTGAAGATGAGGAAAGGTAAGAGAGTGATAAAAAA
 GATAGAGGAGAAAAATAAGAGACAAGTGACATTCGCAAAGAGAAAGAGTCTAATCAA
 GAAGGCATATGAACCTCTCTGTTCTCTGCGATGCCACCTTGGTCTCATCATCTTCTCTCA
 CTCCAACAGGCTCTACGATTTCTGCTCCAACCTCTACCAGCATGGAGAATCTCATCATGAG
 ATACCAAAAGGAAAAAGAAGGTCAAACCACTGCAGAACACAGTTTCCACTCGGATCAGTG
 TTCAGATTGCGTGAAGACGAAGGAATCAATGATGAGAGAGATAGAGAATCTTAAGCTGAA
 TCTTCAATTGTACGACGGACATGGCTTGAATCTCTTGACCTACGACGAGCTCCTTTCTTT
 TGAGCTCCATCTCGAATCTTCTCTACAACATGCTCGAGCTCGCAAGTCTGAGTTCATGCA
 TCAGCAGCAGCAGCAACAAGCATCAAAGCTTAAGGGAAGAAAGGGTCAAGGAAG
 CTCTTGGGAGCAGCTGATGTGGCAAGCAGAGAGACAGATGATGACGTGTCAAAGACAAAA
 AGATCCTGCGCCGGCGAATGAAGGAGGAGTTCCTTTTTTACGGTGGGGAACAACCCACCG
 ACGTCTTCCACCTCCTTAAGCTACCACAACCAGGCCCAAATACAGGCCCATAACTTCTCT
 CTATCTATAAAAAACAACGTATAGTAAAAAGTATTGACCCGGTTTGGTTCGGTTATGTTG
 ATACCAGACTATTAATTAACCTTCGGTTAGACGTATTTACGACTTGATGCTATCTAGACCT
 TTTTGCCCTTCAAAAAA

>G917 Amino Acid Sequence (conserved domain in AA coordinates:2-57)

MRKGKRVIIKKIEEKIKRQVTFAKRKSLIKKAYELSVLCDVHLGLIIFSHSNRLYDFCSN
 STSMENLIMRYQKEKEGQTAEHSFHSDDQSCDCVKTKESSMREIENLKLNLQLYDGHGLN
 LLTYDELLSFELHLESSLQHARARKSEFMHQQQQQQTDQKLKGKEKGQSSWEQLMWQAE
 RQMMTCQRQKDPAPANEGGVPLRWGTTTHRRSSPP*

>G921 (116..1024)

CCAAGATCGACTCTTACTTCAATCTCTCTCAACTTTCTTCTCAGCTTACGGGAACCTC
 CACACATATACATCCACAAGAACCCATATCGAAGATTCTACATATATTTACATGGA
 TCAGTACTCATCTCTTTTGGTTCGATCTTCTATAGATCTCACTATTGGCGTTACTCGTAT
 GCGAGTTGAAGAAGATCCACCGACAAGTGCTTTGGTGGAGAATTAACCCGAGTTAGTGC
 TGAGAACAAGAAGCTCTCGGAGATGCTAACTTTGATGTGTGACAACCTACAACGTCTTGAG
 GAAGCAACTTATGGAATATGTTAACAAGAGCAACATAACCGAGAGGGATCAAATCAGCCC
 TCCCAAGAAACGCAATCCCCGGCGAGAGAGGACGCATTTCAGCTGCGCGGTTATTGGCGG
 AGTGTTCGGAGAGTAGCTCAACGGATCAAGATGAGTATTTGTGTAAGAAGCAGAGAGAAGA
 GACTGTTCGTGAAGGAGAAAGTCTCAAGGGTCTATTACAAGACCGAAGCTTCTGACACTAC
 CCTCGTTGTGAAAGATGGGTATCAATGGAGGAAATATGACAGAAAGTGACTAGAGACAA
 TCCATCTCCAAGAGCTTACTTCAAATGTGCTTGTGCTCCAAGCTGTTCTGTCAAAAAGAA
 GGTTCAGAGAAGTGTGGAGGATCAGTCCGTGTTAGTTGCAACTTATGAGGGTGAACACAA
 CCATCCAATGCCATCGCAGATCGATTCAAACAATGGCTTAAACCGCCACATCTCTCATGG
 TGGTTCAGCTTCAACACCCGTTGCAGCAACAGAGAAGTAGCTTGACTGTGCCGGTGAC
 TACCGTAGATATGATTGAATCGAAGAAAGTGACGAGCCCAACGTCAAGAATCGATTTTCC
 CCAAGTTTCAGAAACTTTTGGTGGAGCAAATGGCTTCTTCTTAACCAAAGATCCTAACTT
 TACAGCAGCTTTAGCAGCAGCTGTTACCGGAAAATTGTATCAACAGAATCATACCGAGAA
 ATAGTTTGTAGCTTCAAATCCGTTAGAGTTTTAGATTGGAATTTGTATGAGTAAGAGAA
 AGAGAGTAGATTATAATCCNTTGTGATACTGAAAAA

>G921 Amino Acid Sequence (domain in AA coordinates: 146-203)

MDQYSSSLVDTSLDLTIGVTRMRVEEDPPTSALVEELNVRVSAENKKLSEMLTLMCDNINV
 LRKQLMEYVNKSNITERDQISPPKKRSPAREDAFSCAVIGGVSESSSTDQDEYLCKKQR
 EETVVKEKVSRYVYKTEASDTLVVKDGYQWRKYQKVTRDNPSPRAYFKCACAPSCSVK
 KKVQRSVEDQSVLVATYEGEHNHPMPSQIDSNNGLNRIHSHGGSASTPVAANRRSSLTVP
 VTTVDMIESKKVTSPTSRIDFPQVQKLLVEQMASSSLTKDPNFTAALAAAVTGKLYQQNHT
 EK*

>G922 (1..1449)

ATGGTGGCTATGTTTCAAGAAGATAATGGAACATCTTCTGTAGCTTCATCACCACCTTCAA
 GTCTTCTCAACTATGTCACTCAACAGACCGACTCTCCTCGCTTCTTCATCTCCGTTTCAT

TGTCTCAAAGATCTCAAACCAGAGGAGCGTGGTCTCTACTTAATCCACCTCTTGCTAACT
 TGTGCCAACACGCTGGCTTCAGGTAGCCTCCAAAACGCTAACGCAGCGCTCGAGCAGCTC
 TCTCACCTCGCTTCTCCTGACGGCGACACGATGCAGCGAATCGCTGCTTACTTCACCGAA
 GCGCTTGCTAAACAGAATCCTTAAGTCTTGCCCTGGTCTTTACAAGGCTCTTAACGCAACT
 CAGACAAGAACTAAACAATGTCTCTGAGGAGATTTCATGTTAGAAGACTCTTCTTTGAGATG
 TTCCCGATACTCAAAGTCTCTTACTTGCTCACTAATCGAGCTATACTCGAGGCTATGGAA
 GGAGAGAAGATGGTTTCATGTGATTGATCTCGATGCTTCTGAGCCAGCTCAATGGCTTGCT
 TTGCTTCAAGCTTTTAACCTCTAGGCCTGAAGGTCCACCTCATTTGAGAATCACTGGTGT
 CATCACCAGAAGGAAGTGCTTGAACAAATGGCTCATAGACTCATTGAGGAAGCAGAGAAA
 CTCGATATCCCGTTTCAGTTTAAATCCCGTTGTGAGTAGGTTAGACTGTTTAAATGTAGAA
 CAGTTGCGGGTTAAAACAGGAGAGGCCCTTAGCCGTTAGCTCGGTTCTTCAATTGCATACC
 TTCTTGCCCTCTGATGATGATCTCATGAGAAAGAACTGCGCTTTACGGTTTCAGAACAAAC
 CCTAGTGGAGTTGACTTGCAGAGAGTTCTAATGATGAGCCATGGCTCTGCAGCTGAGGCA
 CGTGAGAATGATATGAGTAACAACAATGGGTATAGCCCTAGCGGTGACTCGGCCCTCATCT
 TTGCCCTTACCAAGTTTCAGGAAGGACTGATAGCTTCTCAATGCTATTTGGGGTTTGTCT
 CCAAAGTTCATGGTGGTCACTGAGCAAGACTCAGACCACAACGGCTCCACACTAATGGAG
 AGGCTATTAGAATCACTTTACACCTACGCAGCATTGTTTGTGCTTGGAAACAAAAGTT
 CCAAGAACGTCTCAAGATAGGATCAAAGTGGAGAAGATGCTCTTCGGGGAGGAGATCAAG
 AACATCATATCCTGCGAGGGATTGAGAGAAGAGAAAGACACGAGAAGCTTGAGAAATGG
 AGCCAGAGGATCGATTGGCTGGTTTGGGAATGTTCTCTTAGCTATTATGCGATGTTG
 CAGGCTAGGAGATTGCTTCAAGGGTGCAGTTTGTATGGGTATAGAATCAAGGAAGAGAGC
 GGGTGGCGAGTAATTTGCTGGCAAGATCGACCTCTATACTCGGTATCAGCTTGGAGATGC
 AGGAAGTGA

>G922 Amino Acid Sequence (conserved domain in AA coordinates:225-242)

MVAMFQEDNGTSSVASPLQVFSTMSLNRPTLLASSSPFHCLKDLKPEERGLYLIIHLLLT
 CANHVASGSLQANAALEQLSHLASPDGDTMORIAAYFTEALANRILKSWPGLYKALNAT
 QTRTNVSEIEIHVRRLPFEMFPILKVSYLLTNRAILEAMEGEKMHVIDLDASEPAQWLA
 LLQAFNSRPEGPPLRLITGVHHQKEVLEQMAHRLIEAEKLDIPFQFNPVVSRLDCLNVE
 QLRVKTGEALAVSSVLQLHTFLASDDDLMRKNCALRFQNNPSGVDLQRVLMMSHGSAAEA
 RENDMSNNNGYSPSGDSASSLPLPSSGRTDSFLNAIWGLSPKVMVTEQSDHNGSTLME
 RLLESLYTYAALFDCLETKVPRTSQDRIKVEKMLFGEEIKNIIISCEGFERRERHEKLEKW
 SQRIDLAGFNVPLSYAYMLQARRLLQCGFDGYRIKEESGCAVICWQDRPLYSVSAWRC
 RK*

>G932 (206..1213)

CCACGCGTCCGACCACCTTGTACCTCTTTGTCTTAAGTACTCTTTAACCTACAATTCCT
 AAGCTCTCAAGCCACAAAAACCAAAACCGTTCTTACCAATATATATATCTGATCATC
 ATCAAAGTCTTCTCTCTGCTCATACCACAAACCGTTCCATTCTTCCCTAATCACAAAG
 TGATATTTACATAGAGAAGATAGAGATGGGAAGACCACCATGCTGTGACAAGATTGGAGT
 GAAGAAAGGACCATGGACACCAGAGGAAGATATCATCTTGGTTTCTTACATCCAAGAACA
 TGGTCTGGAACTGGAGATCTGTGCCTACTCACACAGGTTTGGAGAGATGTAGCAAAAG
 CTGTAGATTGAGGTGGACTAATTATCTTCGACCTGGGATCAAGCGTGGAAATTTACCGA
 GCATGAAGAGAAGATGATTCTCCATCTTCAAGCTCTTTTGGGAACAGGTGGGCAGCTAT
 AGCATCATATCTTCCAGAAAGGACAGACAATGATATAAAGAACTATTGGAACACTCATTT
 GAAGAAAAAGCTCAAGAAGATGAATGATTCTTGTGATAGTACTATCAACAATGGCCTTGA
 TAATAAAGACTTCTCCATATCAAAACAAAACACTACCTCACATCAAAGCAGCAACTCCAG
 TAAAGGTCAATGGGAGAGAAGGCTTCAGACAGATATCAACATGGCTAAACAAGCTCTTTG
 TGATGCCTTGCTATTGACAAACCAAAAACCCAACTAATTTTCTATTCCCGATCTTGG
 TTATGGTCCATCATCTTCTCGTCCTCTACCACCACCACCACCACCACCACCACCAGAG
 AAACACTAATCCATACCCATCTGGGGTCTATGCTTCAAGTGTGAGAACATTGCTCGTTT
 GCTTCAGAATTTTATGAAAGACACACCAAAGACCTCGGTGCCCTTGCCGGTTGCAGCCAC
 CGAGATGGCTATCACACGGCAGCTTCGAGCCCTAGCACAAACGAAGGAGACGGAGAAGG
 GATTGACCATTCTTTGTTTCAGCTTCAACTCCATAGATGAAGCTGAAGAGAAGCCTAAACT
 AATAGACCATGACATTAATGGTCTAATTACACAAGGCTCTCTTTCTTTGTTTCGAGAAATG
 GCTCTTTGATGAGCAAAGCCACGATATGATCATCAATAACATGTCACTAGAGGGTCAGGA
 AGTGTGTTCTAGAAAGCATTAAAGTTTGACGATTTGCTTGGGAACACGAGGCTTAGT
 TATAACAATTTGTATAATTAAAGTACTCTTTAGTTTGTGTTTCAATCTTATTATGATCA
 TATTGCAGTAATTAGGGATTTTAGTCTTTAGTAGTAACCTTAAAGTTTTAACACATTTTT

CTCTATCTTTTTTAGTAGTAACCTCTTTATTTTTTCCTTAAATCTTTGTGCGACGTGGAGATG
ATATCTTCTATGTAGTAGAACTCAAAAGTGACATCATCTTTATTAATGTAACGTCTTT
TTAAAAAAAAAAAAAAAAAAAA

>G932 Amino Acid Sequence (domain in AA coordinates: 12-118)
MGRPPCCDKIGVKKGFWPTPEEDIILVSYIQEHGPGNWRVPTHTGLRRCSKSCRLRWNTNY
LRPGIKRGNFTEHEEKMILHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKLKKMN
DSCDSTINNGLDNKFISNKNVTSHQSSNSSKGQWERRLQTDINMAKQALCDALSIDKP
QNPTNFSIPDLGYGPSSSSSSSTTTTTTTTTTRNTNPYPSPGVYASSAENIARLLQNFMDT
PKTSVPLPVAATEMAITTAASSPSTTEGDGEGIDHSLFSFNSIDEAEKPKLIDHDINGL
ITQGSLSLFEKWLFDQSHDMIINNMSLEGQEVLF*

>G599 (152..1579)

TCGACAGAACAGCTTCGTTGTCACCTTGTCTATTCTATAAATCGCATCCCCATTGACAACCT
TTCACCTTCCATCAAACTCTCTCTCTATATCTCTCTCTATATATCTCTCTCTATATCT
CTCTCTCTCTTCACTCTCTCTTTCTTTTCAAAATGGAAAACTCATGGTCCGACATGGAG
ACCCGACCCGGTTTACCGTCCACCGGAAACACCACTCGAACCGATGGAGTTTTAGCTCG
TTCATGGAGCGTCTCTGCTCTCGAAGTCTCCAAGGCTCTAACACCAACCAACCTCAGAT
TCTCTCTCTCAAAACCGAAGAAGAAGAAGAAGAAGAACCCATCTCTCTGTCTGCTAGACGG
CGACGGCGACACGGAAGACACCGGACTTGTACCGGAAACCCATTCTCTCTCGCTTGTTC
AGAAACTTCTCAAATGGTCATGGATCGTATCTTGTCTCACTCTCAAGAAGTATCACCAAG
AACATCTGGTCGGCTATCTCACAGTAGTGGTCCACTTAATGGTCTTTTGACCGACAGTCC
TCTGTGTCTCTCCCGAATCCGACGACATTAAGCAATTTTGCAGAGCGAACAATAATTC
ATTGAACAGTGTAAATCTCAGTTCGGTTCACGGCGGCAACTCCGGGACCTATAACCGC
TACAGCTACACAGTCCAAGACGGTGGGACGGTGGCTTAAGGACCGGAGAGAGAAAAAGAA
AGAGGAGACTCGGGCTCATAACGCTCAGATTACGCTGCTGTCTGTCTGCGCGCGCTTGC
TGCAGCTGTTGCTGCTATTGTCAGCAGCCACCGCTGCGTCTTCTAGCTGTGGTAAGGATGA
GCAGATGGCTAAACTGACATGGCCGTTGCTTCTGCTGCGACCTTGTGGCTGCTCAGTG
TGTGGAAGCTGCTGAAGTTATGGGAGCTGAGAGAGAGTATTGGCTTCTGTTGTTAGCTC
CGCCGTCAATGTTCTGCTGCGGAGATATTATGACTCTCACCGCCGGAGCAGCTACAGC
TTTAAGAGGAGTGCAAAACATTGAAGGCAAGGGCAATGAAGGAAGTGTGGAACATAGCATC
AGTGATACCAATGGATAAAGGACTCACTTCTACAGGAGGAAGCAGCAATAATGTTAATGG
TAGCAATGGAAGCTCAAGCAGTAGTCAAGTGGTGAACCTGTACAACAGGAGAATTTCTTT
GGGAACCTGTAGTAGAGAATGGCTCGCTAGAGGTTGTGAACCTCTCAAACGCACTCGCAA
AGGTGATCTCCACTGGAAGATAGTATCTGTTTACATCAACAAAATGAATCAGGTTATGTT
GAAGATGAAGAGCAGGCATGTTGGAGGAACCTTCACCAAGAAGAAAAAGAACATTGTGCT
TGATGTGATCAAGAATGTCCCGGCTGGCCTGGACGACATTGCTAGAGGGAGGAGATGA
CTAAGATACTTCGGTTTGAAGACGGTTATGCGAGGTGATGTTGAATTCGAGGTCAAGAG
CCAAAGGGAATATGAATGTGGACACAAGGTGTCTCAAGGCTCTTGTGTTCTGCTGCTGA
GAGGAAGTTTAGGATGTGAATAAACGTTCAATGGCTGCTTGGTTAAGTGTGAGTTTTTTT
TTTAACTTATGTGGTCAAATTTCAATAGTAGGGGTTCTTTTAAAGGTAATGTTTTTTGGG
TTGGGTATAGGATAAAATGGACCTACCAGTCAAGGTGAGGAAGCATTGGGTAAACAAAA
CTTAGTGGGGGTGATCTGTAATATCTATGTTCTTAGTTTTTTTTTGGTTGTGGTGGTCT
TTTTGTATAAAAAACAAAGTTGAAGTAATAGATATATAGTATGTTTTTAATTTTAA

>G599 Amino Acid Sequence (domain in AA coordinates: 187-219, 264-300)

MEKLMVPTWRPDPVYRPPETPLEPMEFLARSWSVSALEVSKALTPPNPQILLSKTEEEEE
EPISSVVDGDDTETDGLVTGNPFSSACSETSQMVMDRILSHSQEVSPRTSGRLSHSSG
PLNGSLTDSPPVSPPEDDIKQFCRANKNSLNSVNSQFRSTAATPGPITATATQSKTVGR
WLKDRREKKKEETRAHNAQIHAASVAVGVAAGVAAIAAATAASSSCGKDEQMAKTDMAVA
SAATLVAAQCVEAAEVMGAEREYLASVVSSAVNVRSAGDIMTLTAGAATALRGVQTLKAR
AMKEVWNIAVIPMDKGLTSTGGSSNNVNGSNGSSSSSHSGELVQQENFLGTCSEWLAR
GCELLKRTRKGLDHLWKIVSVYINKMNQVMLKMKSRHVGGTFTKKKKNIIVLDVIKNVPAWP
GRHLLLEGDDLRYFGLKTVMRGDVEFEVKSQREYEMWTQGVSRLLVLAAERKFRM*

>G804 (114..1139)

ATACTCCAAGAAATTTATAGGTTATAAGTAAAAATTCAGTACAAGTTTGTGTTGTTGTTA
TTCCATTTTCTGTGTGTTTTTTTCCCATTAATTTATAAATTTTATAAGCAATATGGAGT
CCCACAACAACACAGAGCAACAACAACACCACTGGTTCGGCCCATCTGGTCCCATCCA
TGGGACCAATCTCCGGTTCAGTCTCATTAAACCACTGCTCCAACTCCACTACCACCA
CCGTCAACCGCCGCTAAACACCCGCAAAACGACCGTCCAAGGACCGTCACATCAAAGTAG

ACGGACGTGGCCGGAGGATACGTATGCCGGCTATCTGCGCAGCACGTGTCTTCCAACATA
CACGTGAGTTACAACACAAATCGGACGGCGAGACTATAGAGTGGCTGCTCCAACAAGCGG
AGCCAGCTATCATCGCAGCCACCGGAACGGAACCATACCGGCGAATATCTCTACTTTGA
ACATCTCTCTTCGAAGCAGTGGCTCTACTCTTTCAGCTCCACTGTCTAAATCTTTCCACA
TGGGAAGAGCGGCTCAAAACGCTGCCGTTTTTGGGTTCCAGCAACAGCTTTATCATCCTC
ATCATATCACGACAGATTCTTCTTCTTCTCTTCCCAAACATTCCGTGAAGAAGATC
TTTTTAAAGATCCTAATTTTCTAGATCAAGAACCCGGTTCAAGATCACCTAAACCGGGAT
CCGAAGCTCCTGATCAAGATCCGGGTTCCGACCCGGTCAAGAACACAAATATGATACCGC
CGATGTGGGCACTAGCGCCAACGCCAGCCTCCACAAACGGAGGTAGTGCTTTTGGATGT
TACCAGTCGGAGGAGGAGGAGGTCCGGCTAACGTTCAAGATCCATCACAGCACATGTGGG
CGTTTAATCCGGGTCATTACCCGGGTCGAATCCGGGTCGGTTCAGCTAGGGTCTATGTTAG
TGGGAGGTCAACAGTTAGGGTTAGGTGTTGCAGAAAATAACAATTTGGGGCTATTTTCCG
GCGGAGGAGGACGGTTCGGGTTGGTCTCGGAATGAGTCTTGAGCAAAAGCCTCAAC
ATCAAGTGAGTGATGCTACTAGAGACCAAATCCTACTATAGATGGTTCTCCTTGAA
AGACTTCATGATTTCTTTGGTTTTTAAAAAGTGTGAATGTGTGATTTATTGCAACTTTTG
TTGAGGACTCCAATGTTAATATGGGTTTTAGGGTTGGCTTTTCGGGATTGCCAATTGTT
ATT

>G804 Amino Acid Sequence (domain in AA coordinates: 54-117)
MESHNNQSNNTTGSALVPSMGPISGSVSLTTAPNSTTTTVAAKTPAKRPSKDRHI
KVDGRGRRIRMPAICARVFLTRQLQHKSDGETIEWLLQQAEPALIAATGTGTIPANIS
TLNISLRSSGSTLSAPLSKSFHMGRAAQNAAVFGFQQQLYHPHHITDSSSSSLPKTFRE
EDLFKDPNFLDQEPGSRSPKPGSEAPDQDPGSTRSRTONMIPPMWALAPTPASTNGGSF
WMLPVGGGGGPANVQDPQSQHMWAFNPGHYPRIGSVQLGSMVLGGQQQLGLGVAENNNLGL
FSGGGGDDGRVGLGMSLEQKPQHQVSDHATRDQNPTIDGSP*

>G1062 (297..1781)

CAAAAAAAGTTTCAATTTTGAAGCTCTGAGAAATGAAATCTATCATTCTCTCTC
TATCTCTATCTTCTTTCAGATTTTCGTTCTTCAATTCATGAAATCCTCGTGATTCTAC
TTTAATGCTTCTCTTTTACTTTTCCAAGTCTCTGAATATTCAAAGTATATATCTTTT
GTTTTCAAACCTTTTGCAGAATTGTCTTCAAGCTTCCAAATTTCAAGTTAAAGGTCTCAACT
TTGCAGAATTTTCTCTAAAGGTTTCAAGCTTTGGGGTAAAGGTGTCAACTTTGGCGATGG
GTCTTGACGGAAACAATGGTGGAGGGTTTGGTTAAACGGTGGTGGTGGAGAAAGGGAAG
AGAACGAGGAAGGTTCAATGGGAAGGAATCAAGAAGATGGTCTTCTCAGTTTAAGCCTA
TGCTTGAAGGTGATTGGTTTAGTAGTAACCAACCACATCCACAAGATCTTCAGATGTTAC
AGAATCAGCCAGATTTTCAGATACTTTGGTGGTTTTCTTTTAAACCTAATGATAATCTTC
TTCTTCAACACTCTATTGATTCTTCTTCTTCTTGTCTCTCTTCTCAAGCTTTTAGTCTTG
ACCTTCTCAGCAAAATCAGTTCTTGTCAACTAACAACAACAAGGGTGTCTTCTCAATG
TTCTTCTTCTGCAAAACCTTTTGATAATGCTTTTGAGTTTGGCTCTGAATCTGGTTTC
TTAACCAATCCATGCTCCTATTTTCGATGGGGTTTGGTTCTTTGACACAATTGGGGAACA
GGGATTTGAGTTCTGTTCTGATTTCTTGTCTGCTCGGTCACTTCTTGCGCCGGAAAGCA
ACAACAACAACAATGTTGTGTGGTGGTTTTCAGAGTCCGTTGGAGTTGGAAGGTTTGG
GTAGTCTCTGCTAATGGTGGTTTTGTTGGGAACAGAGCGAAAGTTCTGAAGCCTTTAGAGG
TGTTAGCATCGTCTGGTGACAGCCTACTCTGTTCCAGAAACGTGCAGCTATGCGTCAGA
GCTCTGGAAGCAAAATGGGAAATTCCGAGAGTTCGGGAATGAGGAGGTTTAGTGATGATG
GAGATATGGATGAGACTGGGATTGAGGTTTCTGGGTTGAATATGATCTGATGAGATAA
ATGAGAGCGGTAAAGCGGCTGAGAGTGTTCAGATTGGAGGAGGAGGAAAGGTAAGAAGA
AAGGTATGCCTGCTAAGAATCTGATGGCTGAGAGGAGAAGGAGGAAGAAGCTTAATGATA
GGCTTTATATGCTTAGATCAGTTGTCCCCAAGATCAGCAAAATGGATAGAGCATCAATAC
TTGGAGATGCAATGATTATCTGAAGGAACCTTCTACAAAGGATCAATGATCTTCACAATG
AACTTGAGTCAACTCCTCTGGATCTTTGCTCCTCAACTTCATCAAGCTTCCATCCGTTGA
CACCTACACCGCAAACTCTTCTTGTCTGTCAAGGAAGAGTTGTGTCCCTCTTCTTTAC
CAAGTCTTAAAGCCAGCAAGCTAGAGTTGAGGTTAGATTAAAGGAAGGAAGAGCAGTG
ACATTCATATGTTCTGTGGTCTGATACCGGCTGTTGCTGCTACCATGAAAGCTTTGG
ATAATCTTGGATTGGATGTTTCAGCAAGCTGTGATCAGCTGTTTAAATGGGTTTGCCTTGG
ATGTTTTCCGCGCTGAGCAATGCCAAGAAGGACAAGAGATACTGCCTGATCAAATCAAAG
CAGTGCTTTTTCGATACAGCAGGATGATGCTGATGATGATCTGATCTGATCCTGACTTCGAGT
CCATTAAGCATCTGTTGAAGCAGAGCTAGAAGAATAAGTCCCTTTAAATCTGCAATTTT
CTTCTCAACTTTTTTTCTTATGTATAACTTCAATCTAAGCATGTAATGCAATTGCAAT

>G1521 (1..891)

ATGCCTCCATTACCGTCTCCACGGCGCCTTCGTCTTCGAGACATCTTCGATCGCCGGAA
AGTATCGCGAAATTTGCAGGGAGAGCAATATTTCTGCTTTACAGGGGAAATCGTGTCG
ATATGCCTCGAAAACTAACCAGAGCGAAGATCCGCCGCCGTGATCACGGTGTGCAAGCAC
GGATACTGCCTTGCTTGATTTCGGAAGTGGAGCAGCTTCAAGAGGAATTGTCTCTTTGT
AACACTCGTTTTGATTCTGGTTTATCGTTAGTGATTTTGCTTCTAGAAAATACCATAAG
GAGCAATTACCAATTCTTCGTGATCGTGAGACTTTAACTTATCATCGGAATAATCCTTCC
GATCGCCGGAGGATAAATTCAAAGGTTCGAGGGATGTTTTGGAAAACCTCTAGCTCAAGATCA
AGGCCATTGCCATGGCGGAGATCATTTGGACGACCAGGTTCACTTCTGATTCTGTTATC
TTCCAGCGAAAGCTTCAGTGGCGAGCTAGCATATACACTAAGCAATTACGAGCTGTTTCA
TTACATTCAAGGCGCTTGGAACTAAGTTTGGCGGTGAATGATTACACCAAGCAAAGATA
ACTGAAAGAATTGAGCCATGGATTAGAAGAGAGCTTCAGGCAGTCTTGGAGATCCTGAT
CCCTCAGTTATTGTTTCATTTTGGCTCAGCTCTTTTCATCAAAAGGCTTGAGAGAGAGAAT
AATCGACAAACCGGGCAGACCGGGATGTTGGTGGAGATGAAGTCTCCTCTCTTCGAAAA
TTCTTGTCTGATAAGTTGGATATATTTTGGCATGAACCTAAGATGTTTTGCGGAGAGTATA
CTCACGATGGAGACTTATGATGCAGTGGTTGAATACAATGAGGTGGAGTAA

>G1521 Amino Acid Sequence (domain in AA coordinates: 39-80)
MPPLPSSTAPSSSRHLRSPESIAKFAGRAIFPALQKSCPICLENLTERRSAAVITVCKH
GYCLACIRKWSFKRNCPLCNTRFDSWFIVSDFASRKYHKEQLPILRDRETLTYHRNNPS
DRRRIIQSRDVLNENSSSRPLPWRRSFGRPGSVDPVSIVFQRLQWRASIYTKQLRAVR
LHSRRLSLAVNDYTKAKITERIEPWIRRELQAVLGDPPSVIVHFASALFIKRLEREN
NRQTGQTGMLVEDEVSLRKLFLSDKVDIFWHELRCFAESILTMETYDAVVEYNEVE*

>G183 (1..1458)
ATGAGTGATTTTGATGAAAACCTTCATCGAAATGACGTCGTATTGGGCTCCACCATCCAGT
CCTAGCCCAAGAACGATATTGGCAATGCTGGAGCAAACCGACAATGGTCTGAATCCAATC
AGTGAGATCTTCCCTCAAGAAAGCTTGCCAAGAGATCATACTGATCAATCTGGACAAAGA
TCTGGTCTTCGTGAGAGACTGGCTGCAAGAGTAGGATTCAATCTTCCAACACTCAATACA
GAAGAAAACATGAGTCTTGGATGCATTTTTCAGGAGCTCGAATGTTCTTAATCTCCT
GTCGTTGCAATCTCTCCAGGATTCAAGTCCATCAGCACTATTGCATCTCCCAATATGGTC
AGTGATTCTTCCAGATTATCCCTCCGTCTTCAGCCACCAATTACGGACCTCTAGAGATG
GTGGAACCTCCGGTGAAGACAATGCAGCGATGATGATGTTCAACAACGATCTTCTTAT
CAGCCGTACAATGTTGATCTGCCTTCTCTAGAAGTCTTTGATGATATTGCAACGGAAGAG
TCCTTTTATATCCCATCTTATGAACCTCATGTTGACCAATTGGAACCTCTTTAGTCACA
TCCTTTGAATCTGAACCTGTTGACGATGCCCATACCGACATCATCTCCATTGAGGACAGT
GAGACGAGGATGGAACAAAGATGATGACGACGAGGACTTCCAATACGAAGACGAAGAC
GAAGACCAATACGACCAAGATCAAGATGTAGATGAAGATGAAGAGGAAGAAAAAGATGAA
GACAATGTTGCATTAGATGATCCTCAACCTCCACCTCCAAAGAGAAGGAGATATGAGGTA
TCAAACATGATTGGAGCCACAAGAACAAGCAAGACACAAAGGATCATACTTCAGATGGAA
AGCGACGAAGACAATCCTAACGATGTTATCGCTGGAGAAAATACGGTCAGAAAGTCGTC
AAAGGAAATCCTAATCCGAGGAGTTACTTCAAGTGCACAAACATCGAGTGACAGTGAAA
AAACATGTGGAGAGAGGAGCAGACAATATCAAGTTGGTTGTGACTACATACGATGGGATA
CACAACCATCTTCACCACCTGCACGTAGAAGCAATTCCAGTTCAAGGAACCGGTCTGCA
GGGGCAACAATACCTCAAATCAGAATGATCGAACCAGTCGGTTAGGTAGGGCTCCTCCT
ACTCCTACTCCTCCTACTCCTCCTCCTCGTCTTACACACCTGAGGAGATGAGGCCTTTC
TCTTCGTTGGCTACAGAAATTGATCTGACAGAGGTTTATATGACCGGAATCTCTATGCTG
CCGAATATACCGGTTTACGAGAATTCGGGTTTATGTACCAGAATGATGAACCGACGATG
AATGCGATCCCGATGGTTTCAGATGTGTACGATGGGATCATGGAACGCCTGTATTTAAG
TTTGGTGTGACATGTAG

>G183 Amino Acid Sequence (domain in AA coordinates: TBD)
MSDFDENFIEMTSYWAPSSPSRPTILAMLEQTDNGLNPISIEIFPQESLPRDHTDQSGQR
SGLRERLAARVGFNLPPTLNTEENMSPLDAFFRSSNVNPNPVVAISPFGSPSALLHTPNMV
SDSSQIIPSSATNYGPLEMVETSGEDNAAMMMFNNDLPYQPNVDLPSLEVFDIATEE
SFYIPSYEPHVDPIGTPLVTSFESELVDDAHTDIISIEDSESEDGNKDDDEDQFYEDD
EDQYDQDQDVEDEEEKEEDNVALDDPQPPPPKRRRYEVSNMIGARTSKTQRIILQME
SDEDNPNDDGYRWRKYQKVVKGNPNPRSYFKCTNIECRVKKHVERGADNIKLVVTTYDGI
HNHPSPPARRSNSSSRNRSAGATIPQNQNDRTSRLGRAPPTPTPTPTPPSSYTPPEMRPF
SSLATEIDLTEVYMTGISMLPNIPVYENSGFMYQNDEPTMNAMPDGSVDYDGMERLYFK
FGVDM*

>G2555 (177..956)

CTGTTTTTGTATCCGTGTAAATTAATCACACGGTAGTTTTTGATGAAAAGACAACAATCG
GAGAACAATCTGGTCTGCTGCTAAAAATTTAATAAATGTTTTGTCTAATTGTCTCCACCC
ATAAAAAAGCGCGAATTCATTCACCGACTAAAGACATTCTCCGGTGGAGACCCCGATGC
AATCCACTCATATAAGCGGCGGAAGTAGCGGTGGTGGTGGTGGAGGAGGAGGAGGTGA
GTCGAAGTGGATTATCTCGGATCCGTTCCAGCTCAGCTACTTGGATTGAAACCCTACTCG
AAGAAGATGAAGAAGAAGGTTTAAACCTAACCTTTGTTTAAACAGAGCTGCTTACTGGTA
ATAATAACTCTGGAGGAGTGATAACGAGTCGTGACGACTCGTTCCAGTTCCTGAGTTCTG
TTGAGCAAGGATTGTATAATCATCATCAAGGTGGTGGCTTTCACCGTCAGAATAGTTCTC
CGGCTGATTTTCTTAGTGGGCTCGGTTCTGGGACTGATGGGTATTTCTCTAATTTTGGTA
TTCCGGCGAATTATGACTATTTGTGCGACCAACGTTGATATTTCTCCGACTAAACGGTCTA
GAGATATGGAACACAGTTTTCTTCTCAGCTGAAAGAAGAGCAAATGAGTGGTGGGATAT
CAGGAATGATGGATATGAACATGGACAAGATTTTTGAGGATTGAGTTCTTGTAGGGTTC
GTGCTAAACGTGGTTGTGCTACTCATCTCGTAGCATTGCTGAACGGGTGAGAAGAACGC
GAATAAGTGATCGGATTAGGAGGCTGCAAGAGCTTGTTCCTAACATGGATAAGCAAACCA
ACACTGCAGACATGTTGGAGAAGCTGTGGAGTATGTGAAGGCTCTTCAAAGCCAGATCC
AGGAATTGACAGAGCAGCAGAAGAGATGCAAATGCAAACCTAAAGAAGAACAATAATGTA
TCCTTTAGGATTTGATATATCTGTATTTTATTTTGTACTATCTAAAAATGGTGATGATC
TGTTGCAAAATTCGAAACATGATCTTATATATTGAACTAGAAAAAATAGATATATATGAA
TTTTAGCTGTAAATTTTTGTACAATAAGGAGAGAAAAAGATTTAGAAGAGTCAATAAAAAAG
ATGATGTTTACAAGTCAAAAAA

>G2555 Amino Acid Sequence (domain in AA coordinates: 175-245)

MQSTHISGGSSGGGGGGGVSRLSRLSAPATWIELEDEEEGLKPNLCLTELLT
GNNNSGGVITSRDDSFELSSVEQGLYNHHQGGGFHRQNSSPADFLSGSGSGTDGYFSNF
GIPANYDYLSTNVDISPTKRSRDMETQFSSQLKEEQMSGGISGMMDMNMDKIFEDSVPCR
VRKRGKATHPRSAERVRRTIRSDIRRLQELVPMNDKQINTADMLEEAVEYVKALQSQ
IQELTEQQKRCKCKPKEEQ*

>G375 (53..1171)

TCGACAAAACTCTCACTCTCCCTCAAACATAACAAACATACAGAACACAAAATGGGTCT
CACTTCTCTTCAAGTTTGCATGGATTCTGATTGGCTCCAGGAATCCGAGTCATCAGGAGG
AAGCATGTTAGACTCTTCAACGAATTCCTCCGTCAGCAGCCGACATACTAGCAGCTTGCAG
CACTAGACCACAAGCCTCGGCCGTGGCTGTAGCCGCTGCAGCTCTGATGGACGGTGGAG
GAGGCTGCGTCCACCTCAGACCATCTCAAAGTGTCTCGTTGCGAGTCAACACATAC
TAAGTTCTGTACTACAATAACTACAGCCTCTCTCAGCCTCGTTACTTCTGCAAGACTTG
TCGCCGTTACTGGACAAAAGGCGGAACCTAAGGAATATTCCGGTTGGTGGTGGATGCCG
TAAAAACAAGAAACCATCTTCTCTAATTCCTCCTCCTCCACTTCTTCCGGCAAAAAACC
ATCCAACATCGTTACCGCCAATACCTCTGATCTTATGGCTTTAGCACATTCTCATCAAAA
TTACCAACATTCTCCTCTAGGGTTTTCAATTTTGGTGGGATGATGGGGTCTTACTCAAC
TCCGGAGCATGGTAACGTTGGTTTCTTGGAGAGCAAGTATGGCGGTTTGCTTTTCGAGAG
CCCTAGACCTATTGATTTCTTGGACAGTAAGTTTGATCTCATGGGAGTGAAATGACAA
CCTGGTCATGGTTAATCATGGAAGTAACGAGATCATCATCATCATATAATCATCACAT
GGGTCTGAATCAGGTGTAGGTCTTAACAACAACAACAATGGTGGATTAAATGGGAT
TTCTACGGGAGGCAATGGAAATGGTGGTGGTCTCATGGATATATCGACATGCCAAAGACT
TATGCTATCTAATTATGATCATCACCATTACAATCATCAAGAAGATCATCAAAGGGTAGC
AACAATAATGGATGTGAAGCCAAATCCGAAGTTGTTATCGCTTGATTGGCAGCAAGATCA
ATGCTACTCCAATGGTGGTGGTAGCGGAGGCGCAGGAAAATCCGACGGTGGTGGATACGG
CAATGGTGGTTATATCAACGGTTTAGGTTTCGTGGAATGGTTTATGAATGGCTATGG
AACGTCCACTAAAAAATACTCCTTGGTTTGATAAGTTAATCAGAACCTCTTTTTTCTTGT
CGTCATCAACTAGTAGTAGTAATAGTAGTTGGAGACTAGAGAAGCACTTCAAATTAT
TTATGGGTTTGGTTTGCTAAGCCAGTTTTAC

>G375 Amino Acid Sequence (domain in AA coordinates: 75-103)

MGLTSLQVCMDSWLQSESSSGSMLDSSNPSAADILAACSTRPQASAVAVAAAAALMD
GGRRLRPPHDHPQKPCRESTHTKFCYNNYSLSQPRYFCKTCRRYWTGKGTLRNIPVGG
GCRKNKPPSSNSSSTSSGKKPSNIVTANTSDLALAHSHQNYQHSPLGFSHFGMMGS
YSTPEHGNVGFLESKYGGLLSQSPRPIDFLDSKFDLMGVNNDLVMVNHGNSNGDHHHHHN
HHMGLNHGVLNNNNNGGFNGISTGGNGNGGGLMDISTCQRLMLSNDHHHYNHQBHDH
RVATIMDVKNPKLLSLDWQQDQCYSNGGSGGAGKSDGGGYNGGYINGLSSWNGLMN

CACAAACCACAGTCTCTCTTTCTCTCTATCTATCTTCTCTTTCTCTCTATCTCTAT
 CACTGAAACCCAAAGAGATCCACCATTGTCTTTTTTCCTTCACACAGAGAACTGTTT

CTTCCACACTTCCTTTTACTAGGCAGTGTTAACCAATTGAGAGAGAAAAATGATGGTTG
 ATGAAAATGTGGAAACCAAGGCCTCTACTTTAGTGGCAAGTGTGATCATGGGTTTGGAT
 CCGGGTCGGGTCATGATCATCATGGGTTATCGGCGTCTGTGCCTCTTCTGGTGTAACT
 GGAAGAAGAGAAGGATGCCTAGACAGAGACGATCTTCTTCTTCTTAACTTCTCTCTT
 TCCCTCCTCTATGCCTCTATTTCCACGTGCCAATCCTCTCCCGCACGTAAATTG
 ACCCAAGAAAGCTAAGATTCTCTTCCAAAAGGAAGCAAGAACAGTGACGTGAGTCTC
 TCCGACGTATGATACTCCGAAGAAAGCCGCGGAGGCTCACTTGCCGGCACTTGAATGCA
 AGGAAGGGATTCCCTATAAGAATGGAAGATTTGGACGGTTTTCACGTTTGGACCTTCAAGT
 ATAGGTACTGGCCAAACAACAATAGCAGAATGTACGTGCTAGAAAACACAGGCGATTTG
 TGAATGCTCATGGTCTGCAGCTAGGTGACTTCATCATGGTTTACCAAGATCTCTACTCAA
 ACAATTACGTTATACAAGCAAGAAAAGCATCGGAAGAAGAAGTAGACGTAATCAATC
 TTGAAGAAGACGACGTTTACACAAACTTAACAAGGATCGAAAACACTGTGGTTAACGATC
 TTCTCCTCCAAGATTTTAAATCATCACAAACAACAACAACAACAGCAACAGCAACA
 GCAACAAATGTTCTTACTATATCCAGTCATAGATGATGTACCAACAAACAGAGTCTT
 TTGTCTACGACACGACGGCTCTTACCTCCAACGATACTCCTCTCGATTTTGGGTGGAC
 ATACGACGACTACTAATAATTATTACTCCAAGTTCGGAACATTGATGGTTTGGGCTCCG
 TTGAGAATATCTCTCTCGATGACTTCTACTAGATAATCAATCGATGGGCTCATGGTATC
 TTGATGGTGATCAGCTATTTAATATCCTTATAATATATATAAGAATTAATGCAATTTGC
 ATATATATTATCAAGTGTGTGAATATAACATTACAGTTTAAAAAAAAAAAAAAAAAAAA
 >G1014 Amino Acid Sequence (domain in AA coordinates: 90-172)

MVDENVETKASTLVASVDHGFSGSGHDHGLSASVPLLGVNWKRRMPQRSSSSPNL
 LSFPPMPPIPIHVPTPLPARKIDPRKLRFLFQKELKNSDVSSLRRMILPKKAAEAHLPAL
 ECKEGIPIRMEDLDGFHVWTFKYRYWPNNSRMYVLENTGDFVNAHGLQLGDFIMVYQDL
 YSNYVIQARKASEEEVDVINLEDDVYTNLRIENTTVNDLLQDFNHHNNNNNNNSN
 SNSNKCSYYYPVIDDVTTNTESFVYDTTALTSNDTPLDFLGHTTTTNNYYSKFGTFDGL
 GSVENISLDDFY*

>G1035 (103..624)

CCATAATAATATATTAAACCTATATACTATAATCTTTTTACATAATAAATTTGGGTCCT
 GCGTCTTAATCATAGTACTTAATTTTCTCTGTGTGTTTTAATATGAATAATAAACTGAA
 ATGGGATCTTCCACAAGTGGAATTTGCTCGTCCGTTTCAACCACTGGTTTAGCTAACTCC
 GGTTCAGAATCTGATCTCCGGCAACGTGATCTAATCGACGAGCGGAAGAGAAAGAGGAAA
 CAGTCGAACAGAGAATCTGCGAGGAGTTCGAGGATGAGGAAGCAGAAGCATTGATGAT
 CTCACTGCTCAGGTGACTCATCTACGTAAAGAAAACGCTCAGATCGTCCCGGAATCGCC
 GTCACGACGCAGCACTACGTCACTATCGAGGCGGAGAACGACATTCTCAGAGCTCAGGTT
 CTTGAACTTAACACCGCTCTCCAATCTCTTAACGAGATCGTTGATTTCTCGTCAATCTTCT
 TCTTCAGGATTCGATGAGACCGGTCAGGGATTATTTCGACGGTGGATTATTTCGACGGC
 GTGATGAATCCTATGAATCTAGGGTTTTATAATCAACCAATCATGGCTTCTGCTTCTACT
 GCTGGTGATGTTTTCAACTGTTAGAAAACCTTCACATCATTATCATCGTGAGTGAGACTAA
 TCATCGCAGCAGGGGTAAACCTGTAATTTTTCTTATAAATTATGTGATGATGCTTTGTTT
 CTTTATTTTATAAGATGGTTAATTAGTGTTTAAACTGATTGTAATGATAGACAGTGTA
 GAAATGTGTGATATCATGGAGATGGTGATGTGAGTTGGTACAAATATTTAAGATCTTT
 TCTTTCTATATATTAAAGTGAAGAAATAATATTTTGTCAATTTCTTAAAAAAAAAAAAA
 AAA

>G1035 Amino Acid Sequence (domain in AA coordinates: 39-91)
 MNKITEMGSSTSGNCSSVSTTGLANSSESDDLRLQDLIDERKRRKQSNRESARRSRMRK
 QKHLDDLTAQVTHLRKENAQIVAGIAVTTQHYVTIEAENDILRAQVLELNHRLQSLNEIV
 DFVESSSSGFGMETGQGLFDGGLFDGVMNPMNLGFYNQPIMASASTAGDVFNCA*

>G1046 (1..567)-

ATGATTAGACATCTAAAACCTACATGGAGTCGTCTAGTGTCCATCGCTCTCATTGTTTC
 GATATTCTTGATGGAGTCCCACTACACGACGATCAATTTCAACTCGGCATTCTTACCAAAC
 ACTGACTTTAATGTTTCATTTGCAGTCAAACGTATCGACCCGCATCAACAATCAGTCTCAC
 TTAGACCCAAATGCAGAAAACATTTCCATAACGAAGGTCCTTGCTCCAGAAGAAAGAAGA
 GCAAGAAGAATGGTCTCTAACCGGAATCTGCAAGGAGGTACGATATGCGCAAAAAGAAG
 CAGATCGAAGAGCTGCAACAACAAGTTGAACAACCTCATGATGTTGAATCATCACTTGTCT
 GAGAAAGTCATCAACTTGTGGAAGCAACCATCAGATCCTACAAGAGAACTCACAGCTG
 AAAGAGAAAGTCTCTTCTTTCACTTGCTCATGGCAGATGTGCTATTACCCATGAGAAAT
 GCAGAGAGCAACATCAATGACCGCAATGTGAATTATCTAAGAGGAGAACCATCAAACCGT

CCCACCAACAGTCCCTTTGGTAAGTAA

>G1046 Amino Acid Sequence (conserved domain in AA coordinates:79-138)

MIRHLKPMESSVVRSHCFDILDGVPLHDDHFNSAFLPNTDFNVHLQSNVSTRINNQSH
LDPNAENIFHNEGLAPEERRARMVSNRESARRSRMRKKQIEELQQQVEQLMMLNHHLS
EKVINLLESNHQILQENSQLEKVVSSFHLLMADVLLPMRNAESNINDRNVNVLRGEPNSR
PTNSPFGK*

>G1049 (29..550)

CTAACTTTCTTCCCAAGTAACTTCAAAATGCAGCCGCAACAGACGTTTTTCAGCCTCCA
TAACTACCTAAACTCATCGATACTGCAGTCTCCGTATCCTTCTAATTTCCCGATATCTAC
GCCAATTTCCAACCAACGGTCAAAACCCGTACCTCTCTACGGATTCCAAAGCCCTACAAA
CAATCCACAATCCATGAGCCTAAGCAGCAACAACCTCAACATCAGATGAAGCAGAAGAGCA
GCAGACGAACAACATATAATCAACGAGCGGAAGCAGAGAAGGATGATTTCAAACCGAGA
ATCCGCAAGGAGATCGCGTATGAGGAAGCAAAGACACCTTGACGAGCTTTGGTCACAAGT
GATGTGGTTAAGGATCGAGAATCATCAGTTGCTTGATAAGCTTAACAATCTCTCTGAGTC
TCACGACAAGGTTCTTCAAGAGAATGCTCAGCTTAAAGAAGAAACATTTGAGCTTAAGCA
AGTGATCAGCGATATGCAAATTCAAAGCCCTTTCTCTTGCTTTAGAGACGATATAATCCC
CATTGAATAAAGCATTTTTTCCCGATTCTATTTATGAAAATTTCTTCAAGAGTATGTT
TCTTTGTATGTATATGTTGAGATGTATTTTCAAGGTTTGTATAATGACCCCTTACGACG
ACGTTTTTAGATTGTAGTAAATTTATAAACTAAAGAAGATTAGTGTAAATGAAGAACAAA
TATAA

>G1049 Amino Acid Sequence (domain in AA coordinates 77-132)

MQPQTDFSLHNYLNSILQSPYPSNFPITPPTNGQNPYLLYGFQSPTNPNQSMSSLSS
NNSTSDEAEEQQTNNNIINERKQRRMISNRESARRSRMRKQRLDELWSQVMWLRIENHQ
LLDKLNNLSESHDKVLQENAQLKEETFELKQVISDMQIQSPFSCFRDDIPIE*

>G1069 (89..934)

TTGGAACCCCTAGAGGCCCTTTCAAGCAAATCATCAGGGTAACAATTTCTTGATCTTTCTTT
TTAGCGAATTTCCAGTTTCTGGTCAATCATGGCAAACCCTTGGTGGACGAACCAGAGTGG
TTTAGCGGCATGGTGGACCATTCGGTCTCCTCAGGCCATCACCAAAACCATCACCACCA
AAGTCTTCTTACCAAAGGAGATCTTGGAATAGCCATGAATCAGAGCCAAGACAACGACCA
AGACGAAGAAGATGATCTTAGAGAAGGAGCCGTTGAGGTGGTCAACCGTAGACCAAGAGG
TAGACCACCAGGATCCAAAAACAAACCCAAAGCTCCAATCTTTGTGACAAGAGACAGCCC
CAACGCACCTCCGTAGCCATGTCTTGGAGATCTCCGACGCGAGTGACGTCGCCGACACAAT
CGCTCACTTCTCAAGACGCGAGGCAACGCGCGCTTTGCGTTCTCAGCGGGACAGGCTCAGT
CGCTAACGTCACCTCCGCCAAGCCGCCGACCAGGAGGTGTGGTCTCTCTCCAAGGCAG
GTTTGAAATCTTATCTTTAACCGGTGCTTTCTCCTCGGACCTTCCCAACCGGGTCAAC
CGGTTTAAACGGTTTACTTTAGCCGGGTCCAGGGTCAGGTGCTTGGAGGTAGCGTTGTAGG
CCCCTCTTAGCCATAGGGTCGGTCATGGTGATTGCTGCTACTTTCTCTAACGCTACTTA
TGAGAGATTGCCCATGGAAGAAGAGGAAGACGGTGGCGGCTCAAGACAGATTACCGGAGG
CGGTGACTCACGCCCAGAAATCGGTAGTAACCTGCCTGATCTATCAGGGATGGCCGGGCC
AGGCTACAATATGCCGCCGATCTGATTCCAAATGGGGCTGGTCAGCTAGGGCACGAACC
ATATACATGGGTCCACGCAAGACCACCTTACTGACTCAGTGAGCCATTTCTATATATAAT
GGTCTATATAAAATAAATATATAGATGAATATAAGCAAGCAATTTGAGGTAGTCTATTACA
AAGCTTTTGCTCTGGTTGGAAAAATAAATAAGTATCAAAGCTTTGTTTGTCTTAATGGA
AATATAGAGCTTGGGAAGGTAGAAAGAGACGACATT

>G1069 Amino Acid Sequence (domain in AA coordinates: 67-74)

MANPWWINQSLAGMVDHVSVSGHHQNHHSLLTKGDLGIAMNQSQDNDQDEDDPREG
AVEVNNRRPRGRPPGSKNPKAPIFVTRDSPNALRSHVLEISDGSVDVADTIAHFSRRRQR
GVCVLSGTGSVANVFLRQAAPGGVVSLQGRFEILSLTGAFLPGPSPPGSTGLTVYLAVG
QQQVVGGSVVGPLLAIGSVMVIAATFSNATYERLPMEEEDGGGSRQIHGGGDSPPRIGS
NLPDLSGMAGPGYNMPPHLIPNGAGQLGHEPYTWVHARPPY*

>G1070 (170..1144)

TCGACCAGCTTGGATTTCGTTGTTTCATCATTACTACTCTCTTTCTTCTCTAGCTAGCTA
GTTTTGACAGCAAAATAAGAAGCAAAAAAAGGTCAACTAAAAAAGATCTGTTCTTAGAT
CACTCTCTTCTTCTTTTGTATCCAATTCCACCATTGAATCATAGATCATGGATCCAGT
ACAATCTCATGGATCACAAGCTCTTACCTCCTCCTTTCCACGCAAGAGACTTTCAATT
ACATCTTCAACAACAGCAACAAGAGTCTTCTCCTCACCATCACCAGCAACAAGAAACCA
AACCAGTGGTGACCAACAAGGAGGATCAGGAGGAAACCGACAAATCAAGATGGATCGTGA

AGAGACAAGCGACAACATAGACAACATAGCTAACAAACAGCGGTAGTGAAGGTAAAGACAT
AGATATACACGGTGGTTTCAGGAGAAGGAGGTGGTGGCTCCGGAGGAGATCATCAGATGAC
AAGAAGACCAAGAGGAAGACCAGCGGGATCCAAGAACAACCAAAACCACCGATTATCAT
CACACGGGACAGCGCAACCGCGCTTAGAACCCACGTGATGGAGATCGGAGATGGCTGCGA
CTTAGTCGAAAGCGTTGCCACTTTTGCACGAAGACGCCAACCGCGCTTTGCGTTATGAG
CGGTACTGGAATGTTACTAACGTCACTATACGTGAGCCTGGATCTCATCCTTCTCCTGG
CTCGGTAGTTAGTCTTTCACGGAAGGTTTCGAGATTCTATCTCTCTCAGGATCTTTCTCCC
TCCTCCGGCTCCTCCTACAGCCACCGGATTGAGTGTTCCTCGCTGGAGGACAAGGACA
GGTGGTTGGAGGAAGCGTAGTTGGTCCGTTGTTATGTGCTGGTCTGCTGTTGTCATGGC
TGCGTCTTTTAGCAATGCGGCGTACGAAAGGTTGCCCTTTAGAGGAAGATGAGATGCAGAC
GCCGTTTCATGGCGGAGGAGGAGGAGGATCATTTGGAGTCCGCCCAATGATGGGACAACA
ACTGCAACATCAGCAACAAGCTATGTCAGGTCATCAAGGGTTACCACCTAATCTTCTTGG
TTCGGTTTCAGTTGCAGCAGCAACATGATCAGTCTTATTGGTCAACGGGACGACCACGTA
TTGATCAAATATACACACACTCATAATCGTTGCTAGCTAGCTAACGATGAATCATGAG
TTTAGTGGATATATATATGATTAAAGAGGTTAGCTTATGAACATTAATAAGAGTTTGA
TTCTATCGAGCTTCATTATGTTTGGGTCATCGTTC

>G1070 Amino Acid Sequence (domain in AA coordinates: 98-120)

MDPVQSHGSQSSLPFPHARDFQLHLQQQQEFLHHHQQQRNQTGDGQQGSGGNRQIK
MDREETSNDINIDNINNSGSEKIDIDHGGSGEGGGSGGDHQMTRRPRGRPAGSKNPKP
PIIITRDSANALRTHVMEIGDGLVESVATFARRRQRGVCMVSGTGNVTNVTIRQPGSH
PSPGSVSVLHGRFEILSLSGSFLPPAPPTATGLSVYLAGGQGVVGGSVVGPPLLCAGPV
VMAASFNAAYERLPLEEDEMOTPVHGGGGGSLSPMMGQQLQHQQQAMSGHQGLPP
NLLGSVQLQQQHDQSYWSTGRPPY*

>G1076 (198..1076)

ATTTTAGTCTTCTATAAATTCTCTCAATCCTCTCTCATATCTTTTTTCTTAGTTTAA
TTCAATAAAATAGAAAAAACATATACAAATCTACAGAGAAGAGAAGCTTTATTTTAA
CTTGTGTGTGTGTGTGTGTGTGTATATAATTTTATTTTTTTCAAATTAAATCTCTTCT
TTGCTTTTGATGTGGGCATGGCTGGTCTTGATCTAGGCACAGCTTTTCGTTACGTTAATC
ACCAGCTCCATCGTCCCGATCTCCACCTTCACCACAATTCTCCTCCGATGACGTCACCTC
CCGGAGCCGGGATGGGTCAATTCACCGTCGACGACGAAGACAACAACAACCATCAAG
GTCTTGACTTAGCTCTGGTGGAGGATCAGGAAGCTCTGGAGGAGGAGGAGTACGGCG
GGGGAGGAGACGTCGTTGGTTCGTCGTCACGTCGACGACCCCGGATCCAAGAACAAC
CGAAACCTCCGTAATTATCACGCGGAGAGCGCAAACACTCTAAGAGCTCACATTCTTG
AAGTAACAAACCGCTGCGATGTTTTCGACTGCGTTGCGACTTATGCTCGTCGGAGACAGC
GAGGGATCTGCGTTCTGAGCGGTAGCGGAACGGTCACGAACGTCAGCATACTCAGCCAT
CTGCGGCTGGAGCGGTTGTGACGCTACAAGGAACGTTTCGAGATTCTTCTCTCTCCGGAT
CGTTTTCTCTCCTCCCGCACCTCCCGGAGCAACGAGTTTGACAATTTCTTAGCCGGAG
GACAAGGTCAGGTGGTTGGAGGAAGCGTTTGGGTGAGCTTACGGCGGCTGGACCGGTGA
TTGTGATTGCAGCTTCTTTTACTAATGTTGCTTATGAGAGACTTCCTTTAGAAGAAGATG
AGCAGCAGCAACAGCTTGGAGGAGGATCTAACGGCGGAGGTAATTTGTTTCCGGAGGTGG
CAGCTGGAGGAGGAGGAGGACTTCCGTTCTTTAATTTACCGATGAATATGCAACCAAATG
TGCAACTTCCGGTGGAAGGTTGGCCGGGAATTCCGGTGGAAGAGGTCCTTTCTGATGTG
TATATATTGATAATCATTATATATATACCGCGGAGAAGCTTTTCCGGCGAAGAATTTGC
GAGAGTGAAGAAAGGTTAGAAAAGCTTTTAAATGACTAATGAATTTCAAATTATCATCGT
GATTTCCGACATTGTCTTGTTCATCATGTTAAGCTTAGGTTTATTTTTTGTCTTTGTAG
AATTTTATGTTGAATCCTTTTTTTTTTCTGTGAACTCTATTGTGTTCTGCTGCGAAGG
AAAAAAAATTCTCAAAAAAA

>G1076 Amino Acid Sequence (domain in AA coordinates: 82-89)

MAGLDLGTAFRYVNHQLHRPDLHLHHNSSDDVTPGAGMGHFTVDDNENNNHQLDLAS
GGGSGSGGGGGHGGGDVVGRRPRGRPPGSKNPKPPVITRESANTLRAHILEVTNGC
DVFDCAVATYARRRQRGICVLSSGTVTNVSIQPSAAGAVVTLQGTFEILSLSGSFLPPP
APPGATSLTIFLAGGQGVVGGSVVVELTAAGPVIVIAASFNTVAYERLPLEEDEQQQQL
GGGSGGNNLFEVAAGGGGLPFFNLPMNQPNVQLPVEGWPGNSGGRGPF*

>G1089 (31..2427)

AAGTAAGAGAGCTTCTTAAGGAAGAAGAAGATGGGTTGTGCTCAATCAAAGATCGAGAAC
GAAGAAGCAGTTACTCGTTGCAAAGAACGAAAACAATTGATGAAAGACGCCGTCACGTCT
CGTAACGCTTTCGCCCGCGCTCACTCAGCTTACGCTATGGCTCTTAAAAACACCGGAGCT

GCTCTTTCCGATTACTCTCAGCGCGAGTTTTTAGTCTCTAATCACTCGTCTTCTCCGCA
 GCTGCAGCAATCGCTTCTACTTCTTCTCTTCCCACTGCTATATCTCCTCCTCTTCTCTCT
 TCCACCGCTCCGGTTTCTAATTCAACCGCTTCTTCTTCTCCGCTGCGGTTCTCTCAGCCG
 ATTCTGTACTCTTCTCTCTCTCTCTCTCTCCACCACCGCTTCTCTTCAACGTGCTGCT
 ACTATGCCGGAGATGAACGGTAGATCCGGTGGTGGTCTATGCTGGTAGTGACTCAACGGA
 ATTGAAGAAGATGGAGCCCTAGATAACGATGATGATGACGATGATGATGATGATGACTCT
 GAAATGGAGAATCGTGATCGTTTGATTAGGAAATCGAGAAGCCGTGGAGGTAGTACTAGA
 GGAAATAGGACGACGATTCCGGAGAGAGGAGATCATCTTTCAGGAGGAGAAAGCTCCGCCACCTCCC
 CCTTTGGCGAATTCCGCGGCCAATTCGCCGCCACGTGAGCATCAGCATCAACATCAGCAA
 CAGCAACAACAACCTTCTACGATTACTTCTTCCCTAATGTTGAGAATATGCCTGGAACT
 ACTTTAGAAGATACTCTCCACAACCACAACCACAACCAACCAAGGCCTGTGCCTCCTCAA
 CCACATTCACCAAGTCGTTACTGAGGATGACGAAGATGAGGAGGAGGAAGAGGAGGAAGAG
 GAGGAGGAAGAGGAGACGGTGAATGAACGGAACCACTGGTGGAGGAAGACCGAAGAGA
 GTAGAGGAAGTGACGATTGAATTGGAAAAAGTTACTAATTTGAGAGGGATGAAGAAGAGT
 AAAGGGATAGGGATTCCCGGAGAGAGGAGGAATGCGAATGCCGGTGAAGTGCACGCAT
 TTGGCGAATGTATTCTTGAAGCTTGAAGCTACTAGGCTCCATTACCATTTCTAATTTGAGATAAC
 GATGTTTCTAAGATGCTTGAAGCTACTAGGCTCCATTACCATTTCTAATTTGAGATAAC
 CGAGGACATATTGATCACTCTGCTAGAGTGATGCGTGAATTACATGGAATAGATCATTT
 AGAGGAATACCAATGCTGATGATGGGAAAGATGATGTTGATTGGAAGAGAATGAACT
 CATGCTACTGTTCTTGACAAATGCTAGCATGGGAAAAGAAGCTCTATGACGAAGTCAAG
 GCTGGCGAATCATGAAAAATCGAGTACCAGAAAAAGTTGCTCAATTAAATCGGGTGAAG
 AAACGAGGTGGCCACTTCGGATTCTTAGAGAGAGCTAAAGCAGCAGTAAGTCATTTGCAT
 ACAAGATATATAGTTGATATGCAATCCATGGACTCCACAGTTTCAGAAATCAATCGTCTT
 AGGGATGAACAATACTAACCTAAAGCTCGTTCACCTTGTGAGGCGATGGGGAAGATGTGG
 GAAATGATGCAATACATCATCAAGACAAAGCTGAGATCTCAAAGGTGTTGAGATCTCTA
 GATGTTTTCACAGCGGTGAAAGAAACAAATGATCATCATCAGAACGCACCATCCAGCTC
 TTGGCAGTGGTTCAAGAATGGCACACGCAGTTTTCAGGATGATAGATCATCAGAAAGAA
 TACATAAAGCACTTGGCGGATGGCTAAAGCTAAATCTCATCCCTATCGAAAGCACACTC
 AAGGAGAAAGTATCTTTCGCTTCTCGAGTTCCCAATCCCGCAATCCAAAACTCCTCCAC
 GCTTGGTATGACCGTTTAGACAAAATCCCCGACGAAATGGCTAAAAGTGCCATAATCAAT
 TTCGCAGCGGTTGTAAGCACGATAATGCAGCAGCAAGAAGACGAGATAAGTCTCAGAAAC
 AAATGCCAAGAGACAAGAAAAGAAATGGGAAGAAAAATTAGACAGTTTGAGGATTGGTAC
 CACAAATACATCCAGAAGAGAGGACCGGAGGGGATGAATCCGGATGAAGCGGATAACGAT
 CATAATGATGAGGTGCTGTGAGGCAATTCAATGTAGAACAATAAAGAAGAGGTTGGAA
 GAAGAAGAAGAGCTTACCATAGACAAAGCCATCAAGTTAGAGAGAAGTCACTGGCTAGT
 CTTCGAAGTTCGCTTCCCGAGCTTTTTCAGGCAATGTCCGAGGTTGCGTATTCTATGTTTCG
 GATATGTATAGAGCTATAACGTATGCGAGTAAGCGGCAAGCCAAAGCGAACGGCATCAG
 AAACCTAGCCAGGGACAGAGTTCGTAAGAACTAATGTAAGATCAGAGTAATGTCTTCTTC
 TTCTTTGATCTTGAATATTTAAGCACACATACATAACCGTATAGCTAAATCTTTATC
 ATTGCTTTCTTATATTAAGGTTTTGGCTTTTGTAGAAGGTTTCTTACATATGAGATTCA
 TATAGTGTGATTCTTAAGGAAGTCTGTTGAGTAATAAGAAAGTTGTGTATTGAAA
 TAGAGTTGCATTGTTAATTTTG

>G1089 Amino Acid Sequence (domain in AA coordinates 425-500)

MGCAQSKIENEEAVTRCKERKQLMKDAVTARNAFAAAHSAYAMALKNTGAALSDYSHGEF
 LVSNNHSSSSAAAAIASTSSLPTAISPLPSSTAPVSNSTASSSSAAVPQPIPDTLPPPPP
 PPPLPLQRAATMPENGRSGGGHAGSGLNGIEEDGALDNDDDDDDDDDSEMNDRDLIR
 KRSRSGGSTRGNRTTIEDHHLQEEKAPPPPLANSRPIPPPRQHQQHQQQQQPPFYDYF
 FPNVENMPGTTLEDTPPQPQPQPTRPVPPQPHSPVVTEDDEDEEEEEEEEEETVIER
 KPLVEERPKRVEEVTIELEKVTNLRGMKSKGIGIPGERRGMRMPVTATHLANVFIELDD
 NFLKASESAHDVSKMLEATRLHYHSNFADNRGHIDHSARVMRVTWNRSFRGIPNADDGK
 DDVDLEENETHATVLDKLLAWKLYDEVKAGELMKIEYQKKVAHLNRVKKRGHSDSLE
 RAKAAVSHLHTRYIVDMQSMDSVSEINRLRDEQLYLKLVHLVEAMGKMWEMMQIHQRQ
 AEISKVLRLSDVSVQAVKETNDHHHERTIQLLAVVQEWHTQFCRMIDHQKEYIKALGGWLK
 LNLPIESTLKEKVSSPPRVPNPAIQKLLHAWYDRLDKIPDEMAKSAIINFAAVVSTIMQ
 QQEDEISLRNKCEBTRKELGRKIRQFEDWYHKYIQKRGPEGMNPDEADNDHNDDEVAVRQF
 NVEQIKKRLEEEEEEYHRQSHQVREKSLASLRLPELFQAMSEVAYSCSDMYRAITYAS
 KRQSQSERHQKPSQGS*

>G1093 (1..531)

ATGGGTTATCCGGTGGGGTACACTGAGCTCCTCCTCCCAAGAATCTTCCTTCACTTACTC
TCTCTCTTAGGCTTAATACGAACACTCATAGACACGGGTTTTCGGATATTGGGTCTACCC
GACTTTCTCGAATCCGACCCGGTTTCATCGTCATCGTCATGGCTGGAACCACCGTATATG
TCCACGGCGGCGCATCATACCAAGAAAGCTCATTTTCTTCCCAGTGGCGGCGAGGCTA
GCTGGAGAAATCTTGCCCGTCATCAGATTCTCGGAGCTAACTCGACCCGGATTTCGGATCC
GGATCCGATTGCTGCGCGGTGTGCTCCACGAGTTCGAGAACGATGACGAGATCCGACGG
CTGACGAATTGTCAACACATATTTACCGGAGCTGTTTAGACCGTTGGATGATGGGTAT
AATCAGATGACGTGTCCACTTTGTAGAACCGCTTTATTTCTGATGAGTTACAAGTTGCT
TTTAAACACAGAGTTTGGTCTGAATCTGAACTTCTCGCAGAATCAAATTAG

>G1093 Amino Acid Sequence (domain in AA coordinates: 105-148)

MGYPVGYTELLLPRIFLHLLSLGLIRTLIDTGFRLGLPDFLES DPVSSSSSWLEPPYM
STAAHHHQESSFFFPVAARLAGEILPVIRFSELTRPGFGSGSDCAVCLHEFENDDEIRR
LTNCQHIFHRSCLDWRMMGYNQMTCPLCRTPFISDELQVAFNQRVWSESELLAESN*

>G1127 (191..1351)

GACAGACTCTCTGTATGTGTGCGAGAAGCGAGAAGCGAGAGAGAGAGAGAGAGATTG
TTAGCTCACACGCTTTCTCTATTTTCTCGGAATTCACAAAACAGAAAGTTTCATCCTTTA
CGAGAAATTAAGCCGAAAGAAACAATCTTTGAGTTTGATTTCTTCTTCTTCTCTCTCT
CTCTGCTCTAATGGATTCCAGAGACATCCACCGTCACATAACCAGCTTCAACCACCACC
GGGAATGTTAATGTCTCATTACCGTAACCTTAACGCCGCGCTTACCATTAAATGTTCC
CACTTCCACATCTCAACCGATTCAACACCCTCGTCTTCTTTTGGCAATCAACAACAATC
TCAAACGTTTCATCAGCAGCAACAACAACAATGGATCAGAAGACTCTTGAATCTCTTGG
ATTTGGTGATGGATCACCTTCTTCTCAACCGATGCGATTCCGGATCGATGATCAGAATCA
GCAACTGCAAGTGAAGAAGAAGCGAGGAAGGCCGAGAAAGTATACCTCTGATGGTAGCAT
TGCTTTAGGTTTAGCTCTACGTCTCTCTTCTCTCTGACGTTCTAATTCTTACCGGTGA
GGGTGGTGTGGAGATAGTGGTGGAAATGGAACTCTGTTGATCCACCTGTTAAACGTAA
CAGAGGAAGGCCTCCTGGTCTAGTAAGAAACAGCTTGATGCTTTAGGAGGAACCTCAGG
AGTTGGGTTTACACCTCATGTATTGAAGTGAACACAGGAGAGGACATAGCGTCAAAGGT
GATGGCTTTTTCGGATCAAGGGTCAAGAACAATTTGTATTCTCTCTGCAAGTGGTGCAGT
TTCTAGAGTGATGCTTCGTCAAGCTTCTCATTCTAGTGAATCGTTACTTATGAGGGACG
ATTTGAGATCATTACTCTCTCAGGCTCAGTCTTGAATTATGAGGTAAATGGTTCCACCAA
CAGAAGTGGTAACTTGAGTGTGGCTTTGGCTGGACCTGATGGCGGCATCGTAGGTGGCAG
TGTAAGTTGGTAATCTAGTAGCTGCAACACAAGTCCAGGTGATAGTGGGAAGCTTTGTTGC
AGAAGCAAAGAAACCGAAACAAAGTAGTGTTAACATTGCTCGGGGCAGAATCCTGAACC
GGCTTCAGCCCGCGCTAACATGTTGAACCTTTGGATCAGTCTCTCAAGGACCATCGAGCGA
GTCATCAGAAGAGAATGAGAGCGGTTCTCTGCAATGCACCGTGACAATAATAATGGGAT
ATATGGAGCTCAACAACAACAACAACAACCTCTTCATCCTCATCAGATGCAAAATGTA
CCAACATCTTTGGTCTAATCATGGTCAATAAAATGAAGCGGAATTAATTTGTTCCGTT
TTGGTTACGGTTATGGTTTGATTTCTT

>G1127 Amino Acid Sequence (domain in AA coordinates: 103-110, 155-162)

MDSRDIPPSHNQLQPPPGMLMSHYRNPNAASPLMVPTSTSQPIQHPRLPFGNQQSQT
HQQQQQMDQKTLLESLGFGDGSPSSQPMRFGIDDQNQLQVKKRGRPRKYTPDGSIALG
LAPTSPLLSAASNSYGEAGVSDSGNGNSVDPPVKRNRGRPPGSSKKQLDALGGTSGVGF
TPHVIEVNTGEDIASKVMASDQGSRTICILSASGAVSRVMLRQASHSSGIVTYEGRFEI
ITLGSVNLNYEVNGSTNRSNLSVALAGPDGGIVGGSVVGNLVAATQVQVIVGSFVAEAK
KPKQSSVNIARQNPESAPANMLNFGSVSQGPSSSESEENESGSPAMHRDNNNGIYGA
QQQQQQQLPHQMOMYQHLWSNHGQ*

>G1131 (57..758)

TCGACTCCTCTCTGATTGCTTACCTTCTTCTTTACTACAGGTTTCAGCTCCTCAATGT
CCATGGATTGCTTAAGCTACTTCTTTAACTACGATCCTCCTGTCCAGCTCCAGGATTGCT
TTATTCCCAGATGGATATGATTATCCCTGAAACCGATAGTTTCTTCTTCCAATCTCAAC
CGCAACTGGAGTTTCATCAGCCATTGTTTCAAGAAGAAGCTCCTTCACAGACCCACTTTC
ACCTTTCTGCCACAGTTTCTTCTCCGCAAGAAATCTTCTCCTTACCCCTAAACCTAAACG
AAATCTTCAACGAAACACACGACCTCGATTCTTCTCTCCCAACGCAAAACGCCAGAGAC
TTGTTAACTCCAGCTACAATTGTAACACTCAAAACCATTTCCAGAGCCGTAACCCGAATT
TCTTCGACCTTTTCGGCGACACTGATTTCTCCAGAATCTTGTACCTTCCAGGAGTTTC
GAGTTCCGATTTCTCTTTAGCTTTCAAGGTAGGCCGGGAGATCAAGATGACTCAAAGA

AACCGACGCTTTTCATCTCAGAGCATCGCGGCTAGAGGGAGGAGAAGAAGAATTGCAGAGA
AGACTCACGAGCTCGGAAAACCTATCCCCGGTGGCAATAAACTTAACACCGCCGAGATGT
TCCAAGCCCGCGCTAAGTATGTCAAGTTTTTGAGAGTCAAGTTGGGATTCTCCAACCTGA
TGCAGACCACAAAGAAGGTAATAACCAACCCCAATAAGAACTTTATCATCCAATTGAAA
CTCTAATCGTGTCTTCTCACAAGCTTCTTAATTTGTTTACGCAGGGTAGCTCTAATGTGC
AAATGGAACTCAGTATTTGCTTGAATCGCAAGCAATCCAGGAGAAGTTATCAACAGAGG
AAGTGTGTTTGGTACCGTGTGAAATGGTTCAAGATCTAACAACTGAAGAAACCATTTGCA
GAACCCCGAATATTTCTCGAGAAATCAACAAGTTACTGTCTAAACATCTGGCTAACTAGT
TTTAGTTTCAAGCCTGAAGTTCTCTATGCCTAAATTTGTGTCTGTTATCGTTGTTTTGTC
TTCTTAGTTAGTGTCTTGTCTGTTGATTTAGGGGCTAATTATCCTGGTTAATCTCCTCT
TAACTGGGAA

>G1131 Amino Acid Sequence (domain in AA coordinates: 173-220)
MSMDCLSYFFNYDPPVQLQDCFIPEMDMIIPETDSFFFQSQPQLEFHQPLFQEEAPSQTH
FDPFCDQFLSPQEIFLPNPKNEIFNETHDLDFLPTPKRQRLVNSSYNCTQNHFQSRNP
NFFDPFGDTDFVPESCTCFAGAAVDFSLAFKVGREGDQDDSKKPTLSSQSIAARGRRRIA
EKTHELGKLIIPGKNLNTAEMFQAAKYVKFLQSQVGILQLMQTTKKVITNPK*

>G1145 (243..1142)
GTGATTCTCTCTGCCATTTCTTCGATTGATTCTTGGGTTCTCTTCTCTCGTCTCTC
TTCTGCATGTTTCGCCACTCTACCTTAGAAAAAGGTTACTTTTCGCCTCCGATTTAGGCT
CGATTGATGAATTCGTCGTCGTGGCTATTTATCAAATTGAGCATTAGGGTTTCTGAT
TTGTGGGTTTCAAAATGTTTTATCTATCTGTCTTGTGTTTTTTGTCCGCTACAAAAGC
CTATGGATTCTCAGAGGGGATTGTTGAACAAGCTAAATCTCAGTCTTGAATAGGCAAA
GCTCTCTTTACAGCTTAAACACTTGATGAGGTTCAAATCACTTGGGGAGTTCTGGTAAAG
CTCTGGGAAGCATGAACCTTGATGAGCTTTTGAAGAGTGTCTGTTCTGTTGAAGCTAATC
AGCCATCGTCTATGGCTGTCAATGGTGGAGCAGCTGCTCAGGAGGGTCTTTCTCGCCAGG
GGAGTTTGACTTTGCCTCGGGATCTCAGCAAAAAGACTGTTGATGAGGTTTGGAAAGACA
TTCAGCAGAATAAGAATGGAGGTAGTGCTCATGAGAGGAGGATAAGCAGCCTACACTTG
GGGAAATGACGCTTGAAGACCTGTTGTTGAAGCAGGAGTGGTCACTGAGACTATCCCTG
GTTTGAACCATGATGGTCTGTTGGTGGTGGTAGTGCTGGTTTCAAGTGTGTTTAGGGC
AAAACATTACTCAAGTTGGCCCATGGATTCAATATCATCAGCTCCCATCAATGCCACAGC
CTCAAGCATTATGCCCCTATCCGTTTTCAGATATGCAAGCAATGGTGTCTCAGTCTTCTT
TGATGGGTGGTTTGTGATACACAACTCCTGGAAGGAAGAGGGTAGCTTTCAGGAGAAG
TTGTAGAGAAGACTGTAGAGAGGAGGCAGAAGAGAATGATAAAGAACAGAGAGTCTGCTG
CTCGTTCCCGAGCTAGGAAACAGGCTTACACTCATGAGCTAGAGATCAAAGTTTACCGGT
TAGAAGAAGAAAACGAAAGACTCAGGAAGCAAAAGGAGGTGGAAAAATCCTCCCAAGTGT
ACCACCGCTGATCCCCAAGCGCAGCTCCGACGGACAAGCTCGGCTCCTTTCTGATCTCT
AAACTCTTTTGTCTTTTCTTTTCTTCTCTGTGTCGGTTCACTTATAAAAAAGAGA
GGAAAACAGCTTTGTTTCTTTGTACATTCCGTAGACTTTCTTGACTTGGAGCAATTCTGT
TAACTTTAAAATATTCTCGAGTTATTGTAGTAGCAGACTAGCAGCAGTAATGGTTTTCAT
GAGTCCGATTGAAATTCAGAGATTGAACAGGAAAAAA

>G1145 Amino Acid Sequence (conserved domain in AA coordinates: 227-270)
MDSQRGIVEQAKSQSLNRQSSLYSLTLDEVQNLGSSGKALGSMNLDELLKSVCSVEANQ
PSSMAVNGGAAQEGLSRQGSLLTLPRLDSKKTVDVWKDIQQNKNGGSAHERRDKQPTLG
EMTLEDLLLKAGVVTETIPGSNHDGPVGGGAGSAGLGQNIQVGPWIIQYHQLPSPMPQ
QAFMPYPVSDMQAMVSQSSLMGGLSDTQTPGRKRVASGEVVEKTVERRQKRMKNRESAA
RSRARKQAYTHELEIKVSRLEENERLRKQKEVEKSSQVYHRLIPSGSSDGQARLLSDL*

>G1229 (123..1217)
TTTGGGCGGGTCTTCTTTCCCTAAATCTTTCTTTTATTTTGCTGTTTAAAAAAAATC
CAACCATAAGACAAAACAACGAACGAGGAAGAGAGAGAGAGATATATCTCTAATCA
CGATGCAGGAGATAATACCGGATTCTTCTGAAGAGTGTGAATTTGTGACACTTCACTAG
CCGGAGATGATCTATTGCCATCTTAGAGAGTCTTGAAGGTGCCGAGAGATATCTCCGA
CAGCTGCATCTACACCTAAAGATGGAACCAAGTTCCAAGGAGTTAGTTAAGGATCAAG
ATTATGAAAACCTCATCTCTTAAGAGGAAAAAGCAAAGACTAGAAACCAGGAAAGAAGAGG
ACGAAGAAGAAGAAGACGGAGACGGAGAAGCAGAAGAAGATAATAAGCAAGATGGGCAAC
AAAAGATGTCTCATGTAACCGTGGAACGTAACCGGAGAAAGCAAATGAACGAGCACTTAA
CCGTTTTCGCTTCTTATGCTTGTCTTCTACGTCAAACGGGGGGACCAAGCATCGATCA
TAGGAGGAGTTGTGGAGTACATAAGCGAGTTACAACAAGTTCTCCAATCTTTGGAAGCCA

AGAAACAACGTAAACCTACGCCGAAGTCCTAAGCCCGAGAGTTGTCCCGAGCCCTCGTC
 CTTACCCGCCTGTTCTAAGCCCAAGAAAACCGCCTCTTAGCCCGCGCATCAACCACCACC
 AGATTCACCACCACCTACTTCTCCCTCCCATAAGTCCTCGAACACCTCAGCCAACAAGCC
 CATACCGGGCCATTCCACCGCAACTACCACTCATCCACAGCCTCCGCTTCGCTCTTACA
 GCTCATTTGGCCAGTTGCAGCAGCTTAGGAGATCCACCTCCATACTCTCCTGCTTCATCTT
 CTTTCATCTCCTTCAGTTAGTAGTAACCATGAGAGTAGTGTGATCAATGAGCTTGTGCTA
 ACTCAAAATCGGCTTTGGCTGATGTGGAAGTGAAGTTTTCAGGAGCTAACGTGCTGCTCA
 AAACGGTGTGCGATAAGATCCCGGGACAAGTTATGAAGATAATTGCTGCTCTTGAAGATT
 TGGCTCTTGAGATTCTTCAGGTTAATATTAACACCGTCGACGAAACCATGCTTAATTTCTT
 TCACCATCAAGATTGGAATTGAGTGCCAACTAAGTGCGAAGAAGTGGCTCAACAAATTC
 AGCAACATTCTGCTAGTAAGAAGGATTTAATATAGCTTCGTATAAACCTTAACGAGAG
 AGCAGTAGCTACTCATTCTCTCTCTTAGTATCCCTTTAATTATCTTTTCAGTTTTCGTC
 AAAGATATGGAGTTTAAAAAATAAAATTGTTATCTAAAGTTTAAATCAAAATATTGATTA
 ATTATACTAATATAGGTATAAGTGAGTTTAAAGATTATCAGCTTCATAACAGCCATCG
 TCATGTTTACTTTCTTTTAAATTTTAGAATTTAGACGTACTCCTACCATGTAATTTTATT
 TCTGTCTATACATCAAGCATTGTAGCTGTAATTGCATATGAATGAACAATAGTGTATGAG
 TGATCTCATGAATAATATTCTTCTTGCAACACAAAAA

>G1229 Amino Acid Sequence (domain in AA coordinates: 102-160)
 MQELIPDFLEECEFDVDSLADDDLFAILESLEGAGEISPTAASTPKDGTSSKELVKDQD
 YENSSPKRRKQRLERKEDEEEDGDGEAEDNKQDGGQKMSHVTVERNRRKQMNHLT
 VLRSLMPCFYVKRGDQASIIIGVVVEYISELQQVLQSLQLEAKKQKTYAEVLSRVVPSRP
 SPPVLSPRKPLSPRINHQQIHHHLLLPISPRTPOPTSPYRAIPPQLPLIPQPLRSYS
 SLASCSSLGDDPPYPSPASSSSSPVSSNHSSVINELVANSKSALADVEVKFSGANVLLK
 TVSHKIPGQVMKIIAALEDLALAILQVNINTVDETMLNSFTIKIGIEQLSAEELAQQIQ
 QTFC*

>G1246 (1..1746)
 ATGATCATGTACGGAGGAGGAGGAGCAGGGAAGGACGGTGGATCCACCAATCACTTATCA
 GACGGAGGAGTGATATTGAAGAAAGGTCCATGGACGGCGGCGGAAGATGAGATACTTGCT
 GCGTACGTTAGAGAGAACGGTGAAGGGAATTGGAACGCCGTTTCAGAAAAACACAGGTTG
 GCTCGTTGCGGCAAAAGCTGCCGTCTTCGATGGGCAATCACCTCCGACCAAATCTGAAA
 AAAGGCTCTTTACCGGTGACGAAGAAGCTCTCATCATTACAGCTTCATGCTCAGCTTGGT
 AACAAATGGGCTCGCATGGCTGCTCAGTTACCGGAAGAAGACAGACAACGAGATTAGAAC
 TATTGGAACACGAGATTGAACGACTTCTTCGCCAAGGACTTCCTCTTTATCTCCAGAT
 ATTATCCCTAACCATCAACTCCATCCACATCCACATCATCAACAACAACAGCAACATAAC
 CATCATCATCATCATCAACAACAACAACATCAACAAATGTATTTTCAACCACAA
 TCTTCACAACGAAACACACCATCATCTTCCCCTCTTCCATCTCCAACACCAGCAACGCA
 AAGTCCTCATCATCTTCACTTTTCATACACGACTGTAACCTCCTCCATCCACTTAGC
 CCTCACACTCCAAACACACCATCTCAACTCTCTTCCACACCGCCTCCACCACCACTTCC
 TCTCCTTTATGTTCCCTTCGCAACAACCAATACCCGACCTTCCCCTCTTGGCCCTCCCG
 CGTTCCCAATCAACAACAACAACGGAATTTCACTTCCCCTAGACCTCCACCTCTC
 CTTCACACCGCTTCATCACTCTTCGCAAAACGTTACAACAATGCTAACACTCCTCTTAAT
 TGCATCAACCCGCTCTCAACCGCACCATTTTCCCCTGTTTCAAGAGACTCCTACACTTCC
 TTTCTTACATTGCCTTACCCTTCCCCAACCGCTCAAACCGCTACTTACCACAATACTAAT
 AACCCCTTACTCTCTCTCTCTCTCTTCTTAAACCCCTCTTCTTCTTCTTACCCTACA
 TCAACTTCTTCCCCAAGCTTCTTCACTCCCATTAACCTCTTCTTCCACCTCATTTCAT
 ACCAACCCAGTTTACTCCATGAAACAAGAGCAGCTCCCTTCAAACCAAATCCCCAAATA
 GATGGCTTCAATAACGTCAACAACTTCACAGACAACGAGAGACAGAATCATAACCTTAAC
 AGTTCCGGTGCTCATAGAAGAAGTAGTAGCTGCAGCCTCTTAGAGGATGTCTTGAAGAG
 GCCGAAGCTTTAGCCTCTGGAGGCAGAGGCCGACCTCCAAAACGAAGACAACCTCACAGCT
 TCTCTTCCGAACCACAACAACAACCAACAACAACGACAACCTTCTTCTCGGTTAGTTTC
 GGACATTATGATCTTCTGACAACTTATGTTCTTGAAGATTTGAAATCAAAGGAAGAA
 GAGTCTCTTCAAATGAACACAATGCAGGAGGACATAGCTAAGCTTCTTGATTGGGGAAGT
 GATAGTGGAGAGATCTCTAATGGACAATCATCTGTTGTCACTGACGACAATCTTGTCTT
 GATGTTTCATCAATTAGCTTCACTTATCCCGGCTGATTCTACAGCCGTCGTAGCCGCAACA
 AACGACCAACACAAGAATAATAACAATAATTGTTCTCGGATGACATGCAGGGAATA
 AGGTAG

>G1246 Amino Acid Sequence (domain in AA coordinates: 27-139)

MIMYGGGGAGKDDGGSTNHLSDGGVILKKGPWTAAEDEILAAVYRENGEGNWNNAVQKNTGL
ARCGKSCRLRWANHLRPNLKKGSFTGDEERLI IQLHAQLGNKWMAAQLPGRSTDNEIKN
YWNTRLKRLRLRQGLPLYPPDI IPNHLHPHPPHHQQQQQHNHHHHHHQQQQQHQQMYFQPQ
SSQRNTPSSSPLPSPTPANAKSSSSFTFHTTTANLLHPLSPHTPNTPSQLSSTPPPPPLS
SPLCSPRNNQYPTLPLFALPRSQINNNGNFTFPRPPPLLPSSSLFAKRYNNANTPLN
CINRVSTAPFSPVSRDSYTSFLTLPYPSPTAQATYHNTNPNYSSSPSFLNPSSSSYPT
STSSPSFLHSHYTPSSTSFTNPNVYSMKQEQQLPSNQIPQIDGFNNVNNFTDNERQNHNLN
SSGAHRRSSSCSLLEDVFEEAEALASGGGRPPKRRQLTASLPNHNNTNNDNFFSVSF
GHYDSSDNLCSLQDLKSKEESLQMNMTMQEDIAKLLDWGSDSGEISNGQSSVVTDDNLVL
DVHQLASLFPADSTAVVAATNDQHNKNNNNCSWDDMQGIR*

>G1255 (138..1388)

CAGCTCAAACCTCTCTAGGACTACACTAAATCTAACTTTTTCAGAGAGCAAAGATTCAA
TAATTGAGATTGATCTCAAACCAAAGCTCTCGTGCTCTTGTCGTTGATGTTGGTTGTGT
AGACTTTGTATACAATGATGAAAAGTTTGGCGAATGCTGTTGGAGCGAAGACGGCGAGGG
CTTGCGACAGCTGCGTGAAGAGACGTGCACGGTGGTACTGCGCGGCCGACGATGCTTTTC
TTTGCCAGTCTTGCGACAGTTTGGTCCATTAGCAAACCTCTTGCTCGCCGCCACGAGA
GAGTCCGTTTGAAGACGGCTAGCCCGCGGTCGTAAAGCATAGCAACCACTCATCAGCTT
CTCCTCCACATGAGGTCGCCACGTGGCATCACGGGTTTACTCGTAAAGCTCGAACGCCAC
GTGGCTCTGGTAAGAAAAACAATTCGTCGATATTTTCATGACTTGGTTCTTGATATTAGTA
TTGAGGATCAGACAGACAACTATGAGCTTGAAGAGCAGCTGATCTGTCAAGTGCCGGTTC
TAGATCCGTTGGTGTCTGAGCAGTTCTTGAACGATGTCGTTGAGCCCAAGATCGAGTTTC
CTATGATCAGAAGTGGTTTGTATGATCGAGGAGGAGGAAGACAACGCTGAAAGTTGTCTTA
ATGGATTTTCCCGACCGACATGGAGCTTGAGGAGTTTGTCTGCTGAGCTGGAGACTCTGC
TCGGTTCGGGGTTAGACACGGAGTCGTATGCCATGGAGGAGCTAGGTTTATCTAATTTCAG
AGATGTTCAAAATCGAAAAAGATGAGATTGAAGAAGAAGTAGAAGAGATAAAAGCCATGA
GCATGGATATATTTGATGATGATCGAAAAGACGTGGATGGAACAGTACCGTTTGAGCTAA
GCTTTGATTACGAGTCGTACACAAGACGTCCGAAGAAGAGGTAATGAAGAAGCTTGAAA
GTAGTGGTGAATGTGTGTTAAGGTGAAAGAGGAAGAACATAAGAATGTTCTGATGCTAA
GATTAAACTATGACTCGGTGATATCCACTTGGGGAGGTCAAGGTCCACCGTGGAGTTCAG
GAGAGCCACCGGAACGAGACATGGACATCAGCGGTTGGCCAGCTTTTCCATGGTGGAGA
ATGGAGGAGAAAGTACTCATCAGAAGCAATACGTTGGTGGATGTTTACCATCAAGTGGGT
TTGGAGATGGAGGTAGAGAAGCTAGAGTTTCGAGATACAGAGAGAAGAGGAGGACAAGGT
TGTTTTCTAAGAAGATACGGTACGAGGTACGTAAATTGAATGCAGAGAAAAGACCAGAA
TGAAAGGAAGATTCTGTAAGAGAGCCTCGCTCGCTGCTGCTTACCATTAGGTGTTA
ATTACTGAATAGTTAATATCTATTATCTACTTTTACAAATTTCCGTGAATCT
TTTTTCTTCTGAAACAACAGAGTTATTTTGGCACTTAATTGTGCTTTGAGGACTTGAT
GTACATAGAAGTAACCAATAATAATGTGACTTTTACTA

>G1255 Amino Acid Sequence (domain in aa coordinates: 18-56)

MKSLANAVGAKTARACDSCVKRRARWYCAADDAFLCQSCDSLVSANPLARRHERVRLKT
ASPAVVKHSNHSSASPPHEVATWHHGFTRKARTPRGSGKKNSSIFHDLVPDISIEDQTD
NYELEEQLICQVPVLDPLVSEQFLNDVVEPKIEFPMIRSGLMIEEEDNAESCLNGFFPT
DMELEEFADVETLLGRGLDTSYAMEELGLSNSEMFKIEKDEIEEEVEEIKAMSMDIFD
DDRKDVDGTVPFELSFDYESHKTSEEEVMKNVSSGECVVKVKEEEHKNVLMRLNLYDS
VISTWGGQGPWSSGEPERDMDISGWPAFMSVENGGESTHQKQYVGGCLPSSSGFGDGG
EARVSRVREKRTRLFSKKIRYEVRLNAEKRPKGRFVKRASLAAAASPLGVNY*

>G1304 (1..978)

ATGGGGCGATCACCATTGTTGCGATGAGAATGGTCTAAAGAAAGGGCCATGGACACAAGAG
GAGGATGATAAACTGATAGATCACATTCAAAAACATGGCCATGGCAGCTGGAGAGCTCTT
CCAAAGCAAGCCGGTTTAAACCGATGCGGAAAGAGTTGTAGATTAAGATGGACCAACTAC
TTGAGACCTGACATCAAGAGAGGAAATTTCACTGAAGAGGAAGAACAACCTATTATCAAC
CTCCATTCCCTTCTTGGAACAAGTGGTTCGTCGATAGCCGGTAATCTTCTGGAAGAACG
GACATGAAATAAAAACTATTGGAACACACATTTGAGAAAGAACTTCTCCAAATGGGG
ATTGATCCGGTGACCCATAGGCCAAGAACCGACCATCTAAACGTTTTAGCAGCTCTCCCG
CAGCTTATAGCCGCCGCAAAATTTCAACAGCCTCTTGAATCTCAACCAAAATGTGCAACTG
GATGCAACAACTCTTGCTAAAGCTCAACTGCTACACACTATGATTCAAGTCTTAGCACC
AATAACAACACCACCAATCTTCTTTTCTTCTCATCAACTATGCAAAACAGTAACACCAAT
CTCTTTGGCCAAGCTTCTTACTTAGAGAACCAAAATCTTTTGGTCAGTCTCAAACTTC

TCTCACATTCTTGAGGATGAGAATTTGATGGTCAAAACCCAAATTATTGATAACCCCTTTG
GACTCTTTTTCTTCCCCATACAACCCGGTTTTCAAGATGATCATAATTCCTCCCTCTA
TTGGTTCCGGCGTCTCTGAAGAATCTAAAGAACTCAAAGGATGATCAAGAACAAGAC
ATCGTCGATTACCATCATCATGATGCTTCAAACCTTCATCATCAAACCTCAACGTTTACA
CAAGATCATCATCACCATTGGTGTGACACTATTGATGATGGAGCAAGTGATTCTTTTTGG
AAAGAGATAATAGAGTAA

>G1304 Amino Acid Sequence (conserved domain in AA coordinates:13-118)

MGRSPCCDENGLKKGPTWQEDDKLIDHIQKHGHSWRALPKQAGLNRCGKSCRLRWNTY
LRPDIKRGNTFEEEEQTIINLHSLLGKWKSSIAAGNLPGRTDNEIKNYWNTHLRKKLLQMG
IDPVTHRPRDHLNVLAALPQLIAAANFNLSLLNLNQNVLQDATTLLAKAQLLHTMIQVLST
NNNTTNPSFSSSTMQNSNTNLFQASYLENQNLFQSQNFSHILEDENLMVKTOI IDNPL
DSFSSPIQPGFQDDHNSLPLVLPASPEESKETQRMINKNDIVDYHHHDASNPSSSNSTFT
QDHHHPWCDTIDDGASDSFWKEIIE*

>G1318 (7..849)

AAAAATATGAGGAAGCCAGAGGTAGCCATTGCAGCTAGTACTCACCAAGTAAAGAAGATG
AAGAAGGGACTTTGGTCTCTGAGGAAGACTCAAAGCTGATGCAATACATGTTAAGCAAT
GGACAAGGATGTTGGAGTGATGTTGCGAAAAACCGAGGACTTCAAAGATGTGGCAAAAGC
TGCCGTCTTCGTTGGATCAACTATCTTCGTCCTGACCTCAAGCGTGGCGCTTTCTCTCCT
CAAGAAGAGGATCTCATCATTCGCTTTCATTCATCCTCGGCAACAGGTGGTCTCAGATT
GCAGCAGGATTGCCCTGGTCCGACCGATAACGAGATCAAGAATTTCTGGAACCAACAATA
AAGAAAAGGCTAAAGAAGATGTCCGATACCTCCAACCTTAATCAACAACCTCATCTCATCA
CCCAACACAGCAAGCGATTCTCTTCTAATTCCGCATCTTCTTTGGATATTAAAGACATT
ATAGGAAGCTTCATGTCTTACAAGAACAAGGCTTCGTCAACCCCTTCCTTGACCCACATA
CAAACCAACAATCCATTTCCAACGGGAAACATGATCAGCCACCCGTGCAATGACGATTTT
ACCCCTTATGTAGATGGTATCTATGGAGTAAACGAGGGGTACAAGGGGAACCTTACTTC
CCACCTTTGGAATGTGAAGAAGGTGATTGGTACAATGCAAAATATAACAACCACTTAGAC
GAGTTGAACACTAATGGATCCGGAACGACCTGAGGGTATGAGACCAGTGGAAGAATTT
TGGGACCTTGACCAAGTTGATGAACACTGAGGTTCTTCTGTTTACTTCAACTTCAAACAA
AGCATATGAATATTTTACGTCATCTTATTCTTTTTTCTATTGCGGTTTATACTCAAGAT
TCTTAGCCACACACATAAATGCAAAATATATATACATTGTTAGAGAGTATTTTGTATTT
CGTATAATCTTTTCGTACTAGGGCTTGAGCCTTGAGGTCCCATGTACGATTAGTCAATG
TAAACATATATCTTATAATAAATAAATAAAGAAATAATAAGCACATAAAAAAAAAAAAA
A

>G1318 Amino Acid Sequence (domain in AA coordinates: 20-123)

MRKPEVAIAASTHQVKKMKGLWSPEEDSKLMQYMLSNQGQWSDVAKNAGLQRCGKSCR
LRWNYLRPDLKRGAFSPQEEELIIRFHSILGNRWSQIAARLPGRDNEIKNFWNSTIKK
RLKKMSDTSNLIINSSSSPNTASDSSSNSASSLDIKDIIGSFMSLQEQGFVNPSLTHIQ
NNPFPPTGNMISHPCNDDFTPYVDGIYGVNAGVQGEYFPFLECEEGDWYNANINNHDEL
NTNGSGNAPEGMRPVEEFWDLQLMNTEVPSPFYFNPKQSI*

>G1320 (39..788)

GAAGATCATAAAGATCAAAAGGAGAGAGGTATTAATAATGATGTGTAGTCGAGGCCATT
GGAGACCTGCAGAAGACGAGAAGCTAAGAGAACTCGTCGAGCAATTTGGTCTTCATAATT
GGAACGCCATAGCTCAGAAGCTCTCTGGTCGATCTGGTAAGAGTTGTAGATTGAGATGGT
TTAATCAATTGGATCCTAGGATTAACCGAAACCCCTTCACGGAGGAAGAAGAAGAAAGGC
TTTTAGCGCTCATCGGATCCATGGGAACAGATGGTCTGTGATCGCTAGATTTTTCCTCG
GTGCAACTGATAACGCTGTTAAAAACCATTTGGCACGTCATCATGGCTCGTCTGGCCGAG
AACGGTCCAAGCTCCGTCCACGAGGCTTGGCCATGATGGCACGGTGGCTGCGACTGGGA
TGATTGGTAATTATAAAGACTGCGATAAGGAGAGAAGATTGGCAACCACAACCGCTATCA
ATTTTCTTATCAATTCTCTCATATTAATCATTTTCAAGTCTCAAAGAGTCTTTCACCG
GAAAGATCGGGTTCAGAAATAGTACTACTCAATACAAGAAGGAGCAATAGACCAAACTA
AACGACCGATGGAGTTCTACAAATTTCTCCAAGTAAACCGGATTGGAAGATACACGAAT
TGATAGATAATTCAAGAAAAGACGAAGAAGAAGATGTCGATCAAAACAACCGAATTGTA
ACGAGAATTGTGTTCCATTTTTCGACTTTTGTCTGTTGGAACTCTGCCTCTCAGGGTT
TATGTTAATTTGTCGTACCAATGTAAGGTTGAGCCAAATGTGAAAATTAGTTAAGA
TGTAAGAAAGTACTAATCAATTAGAGCTCTGTTTGGCCAAATGTGAAAATTAGTTAAGA
CATCCCAAACATTTTCTTGTATAACACATATAAGGTTGTACTTTTATCAGGTCTAATTTT
CTATTTTATTTTAAAGGATGTTTAAATCAGACCCATAACCATTCGATAAAAAAAAAAAAAA

>G1320 Amino Acid Sequence (domain in AA coordinates: 5-108)
MMCSRGRHWRPAEDEKLRELVEQFGPHNWNIAIAQKLSGRSGKSCRLRWFNQLDPRINRNPF
TEEEERLLAPHRIHGNRWSVIARFFPGRTDNAVKNHWHVIMARRGRERSKLRPRGLGHD
GTVAATGMIGNYKDCDKERRLATTTAINFPYQFSHINHFOVLKESLTGKIGFRNSTTPIQ
EGAIDQTKRPMEFYNFLQVNTDSKIHLEIDNSRKDEEEDVDQNNRIRNENCVPFFDFLSV
GNSASQGLC*

>G1330 (36..959)

GTACCGGCGACCTCTTTGTGGGTCACTCTTCATCAATGGGTGACAAAGGAAGGAGCTTAA
AGATCAACAAGAACATGGAGGAATTCACGAAAGTGAAGAAGAAATGGACGTAAGGAGAG
GTCCATGGACAGTTGAGGAAGATTTAGAGCTCATCAATTACATTGCTAGTCATGGTGAAG
GTCGATGGAACCTCTCTCGCTCGTTGCGCCGAACCTCAAAGGACCCGGAAGGAGTGCAGAC
TTCCGTGGCTGAACCTCTCCGACCATGTGCGCCGTGGAACATAACCCCTCGAAGAAC
AACTCTTGATTCTTGAACCTTCACACACGTTGGGGCAATAGATGGTCTAAGATTGCACAAT
ATTTACCAGGAAGAACGGATAACGAGATCAAAACTATTGGAGAACACGTGTTCAAAAGC
ATGCAAAACAGCTTAAATGCGACGTGAACAGTCAACAATTTAAAGACACCATGAAGTATC
TTTGGATGCCTCGGCTCGTAGAAAGGATCCAAGCCGCGTCCATCGGGTCTGTTTCCATGT
CATCTTGCGTCACCACTCCTCAGATCAGTTCGTGATCAACAACAACAACACCAACAACG
TGGATAATTTGGCTTAAATGAGTAACCCCTAATGGTTACATCACGCCGGATAATTCCAGCG
TGGCAGTATCTCCTGTATCAGATTGACGGAGTGTCAAGTGAAGTGAAGTGTGAAGA
TTGGTCAGGATGAGAATTTGGTGGATCCAAAATGACATCGCCGAATTATATGGATAATA
GCAGTGGACTATTAAACGGAGATTTTACGAAGATGCAAGATCAAAGTGACCTTAATTGGT
TTGAAATATTAAATGGGATGGTACCAATATTTCGGACAGTTTTTGGAACATTGGAAATG
ATGAAGACTTCTGGCTCTTACAACAACATCAACAAGTCCACGACAATGGAAGCTTCTGAA
TAGACAAGAAGCTATGCGGCC

>G1330 Amino Acid Sequence (domain in AA coordinates: 28-134)
MGDKGRSLKINKNMEFTKVEEMDVRRGPWTVEEDLELINYIASHGEGRWNSLARCAEL
KRTGKSCRLRWLNLYLRPDVRRGNITLEEQLLILELHTRWGNRWSKIAQYLPGRDNEIKN
YWRTRVQKHAKQLKCDVNSQQFKDTMKYLWMPRLVERIQAASIGSVSMSSCVTTSSDQFV
INNNTNNVDNLALMSNPNGYITPDNSSVAVSPVSDLTECQVSSEVWKIGQDENLVDPKM
TSPNYMDNSSGLNGDFTKMQDQSDLNWFENINGMVPNYSDFSFWNIGNDEDFWLLQQHQQ
VHDNGSF*

>G1352 (79..900)

GCGCGATTAAAACTCTCAACTTTTCTCTCAAAATTTCTGATCCTTTGATCCAACAGTTAG
AAGAAGATTTCATCTGATCATGGCCCTCGAAGCGATGAACACTCCAACCTCTTTTACC
AGAATCGAAACGAAAGAAGATTTGATGAACGACGCCGTTTTCATTGAGCCGTGGCTTAA
CGCAAACGCTCCAAACGTCAGCGTTCTCACAGCCCTTCTTCGTCTTCTTCTCACCGCCT
CGATCTCGACCCAAATCCAGAATCAAGATCTTACGGAAGAAGAGTATCTCGCTCTTTGT
CTCCTCATGCTCGCTAAAGATCAACCGTCGCAACGCGATTTTCATCAACAGTCGCAATCG
TTAACGCCGCGCCGAGAAATCAAAGAACCTTCCGTACAAGTGAACGTCGTGAAAAAGCG
TTTCTCTCTATCAGGCTTTAGGCGGTACAAAGCAAGTCACCGAATCAAACCACCAACC
GTAATCTCAACAACCGCCGATGATTCAACAGCTCCGACCATCTCCATCGTCGCCGAGAA
AAACATCCGATTGCTGCCTCCGGAAGATCCACGAGTGTTCATCTGTCATAAAGTGTTT
CCGACGGGTCAAGCTTTAGGCGGTACAAACGTTGTCACTACGAAGGCAACCTCGGCGGC
GGAGGAGGAGGAGGAAGCAAATCAATCAGTCACAGTGAAGCGTGTGAGCACGGTATCG
GAAGAAAGGAGCCACCGTGGATTATCGATCTAAACCTACCGCGCTTACCTGAACCTCAGC
CTTCATCACAATCCAATCGTCGACGAAGAGATCTTGAGTCCGTTGACCGGTAAAAAACCG
CTTTTGTGACCGATCAGCACCAAGTCATCAAGAAAGAAGATTATCTTTAAAAATCTAA
TACTCGACTATTAAFTCTTGTGTGATTTTTTTCGTTACAACCATAGTTTCATTTTCATT
TTTTAGTTACAAATTTTAAATTGTTCTGATTTGGATTGAAA

>G1352 Amino Acid Sequence (domain in AA coordinates: 108-129,167-188)
MALEAMNPTSSFTRIETKEDLMNDVFIIEPWLKRKRQRSHSPSSSSSSPPRSRPKS
QNQDLTEEEYLALCLMLAKDQPSQTRFHQQSLSLTPPPESKNLPYKCNVCEKAFPSYQA
LGGHKASHRIKPTVISTTADDSTAPTISIVAGEKHPAASGKIHECSIHKVFPTGQAL
GGHKRCHYEGNLGGGGGGGSKSISHSGSVSSTVSEERSHRGFIDLNLPALPELSLHNP
VDEEILSPLTGKKPLLLTDHDDQVIKKEDLSLKI*

>G1354 (1..1047)

ATGGAAGTCTCGCACACATTCTCCCGGTTATCGATTCCATCCGACCGATGAAGAATC

GTGACTATTATCTCAAGAACAAAGTTGCATTCCCGGGAATGCAAGTTGATGTTATCAAA
GATGTTGATCTCTACAAAATCGAGCCATGGGACATCCAAGAGTTATGTGGAAGAGGGACA
GGAGAAGAGAGGGAATGGTATTTCTTTAGCCACAAGGACAAGAAATATCCAAGTGGGACA
CGAACCAATAGAGCAACGGGCTCCGGATTTTGGAAAGCAACGGGTCGAGACAAGGCCATT
TACTCAAAGCAAGAGCTTGTGGGATGAGGAAGACTCTTGTCTTTTACAAAGGTAGGGCC
CCAAATGGTCAGAAATCTGATTGGATAATGCACGAATACCGTCTTGAGACCGATGAAAAT
GGACCGCCTCATGAGGAAGGATGGGTGGTTTGTGCGCTTTCAAGAAGAAGCTAACCACG
ATGAACACACAATCCAAGAACAAATGATGGGATCATCATCAGGCCAAGAATCTAACTGG
TTCACGCAGCAAATGGATGTGGGAATGGTAATTACTATCATCTTCTGATCTAGAGAGT
CCGAGAATGTTTCAAGGCTCATCATCATCACTATCATCATTACATCAGAAATGATCAA
GACCCCTTATGGTGTCTGACTCAGCACTATTAACGCAACCCCAACTACAATAATGCAACGA
GATGATGGTTCATGTGATTACCAATGATGATGATCATATGATCATGATGAACACAAGTACT
GGTGATCATCATCAATCAGGATTACTAGTCAATGATGATCATCAATGATCAAGTAATGGAT
TGGCAAACGCTTGACAAGTTTGTGTCTTCTCAGCTAATCATGAGCCAAGAAGAGGAAGAA
GTTAAACAAAGATCCATCAGATAATTCTTGAATGAAACATTTTCATCATCTCTCTGAAGAG
CAAGCTGCAACAATGGTTTCGATGAATGCTTCTTCTCTTCTTCCATGTTCTCTCTAC
TCTTGGGCTCAAAATACACACACGTAA

>G1354 Amino Acid Sequence (domain in AA coordinates: TBD)

MESLAHIPPGRFHPHTDEELVDYLLKNKVAFPQMVDVIKDVLDYKIEPWDIQELCGRGT
GEEREWYFFSHKDKKYPTGTRTNRAATGSGFWKATGRDKAIYSKQELVGMKRLTVFYKGRA
PNGQKSDWIMHEYRLTDENGPPHEEGWVVCRAFKKKLTMMNYPNPTMMGSSSGQESNW
FTQQMDVGNNGNYHLPLDLESPRMFQSSSSSLSLHQNDDQDPYGVVLSTINATPTTIMQR
DDGHVITNDDDHMMNMTSTGDHHSGLLVNDDHNDQVMDWQTLDFVASFQSLMSQEEEE
VNKDPDSNSSNETFHLSEEQAATMVMNASSSSSPCSFYSWAQNTHT*

>G1360 (1..1257)

ATGGGAGATAGAAACAACGACGGTGATCAGAAAATGGAGGATGTATTGTTGCCCGGATTT
AGGTTTCATCCAACCGACGAAGAGCTCGTAAGCTTCTACCTGAAGCGGAAGGTTCAACAC
AACCCTCTCTCCATTGAGCTCATAAGACAACCTCGATATCTACAAATATGACCCCTGGGAT
CTTCCAAAGTTTGCGATGACGGGTGAAAAAGAATGGTACTTTTATTGTCCAAGGGACAGG
AAGTATAGGAACAGCTCGAGGCCAAACCGAGTGACCGGAGCTGGTTTTTGGAAAGCCACG
GGAACGGACCGGCCGATATACTCGTCAGAAGGAACAAATGCATAGGTTTAAAGAAGTCC
TTAGTGTTCTACAAAGGAAGAGCAGCGAAAGGAGTTAAGACTGATTGGATGATGCATGAG
TTTCGTTTGCCTTCTCTCTCCGAACCATCTCTCTCTTCTAAGAGATTCTTCGACTCTCT
GTCTCTCCCAAGATTTCATGGGCTATATGCAGAATCTTCAAAAAGACCAACACAACGACC
CTAAGAGCTCTCTCTCTACTCTTTTGTCTTCTGTTACCAACAGAAACAAGCACCGACACA
ATGTCTAACCAAAAGCAATCAACACATACCATTTTTCTTCAGACAAGATCCTCAAACCT
AGCTCTCACTTCCAGTTTCCCATGAGAAATGAACACTCCCAAACTAGTAATAGTACA
ACTCCATCCGTTCCTACTATAAGTCCCTTCTCTTACTTGGATTTCACTTCATACGACAAA
CCCACCAACGTTTTCAATCCGGTTTCATGTTTAGACCAACAATACCTCACAAATCTCTTT
CTTGCCACACAAGAAACACAACCTCAGTTTCCAGGCTCCCTCGTCAAAATGAAATCCCA
TCGTTTCTGCTAAACACGCTTTCAGATTGACCTTCTTGGGAGAATTACAGAGCCATATC
GACCTCAGCGCAGTGTTGGCCCAAGAGCAATGTCCCCGCTTGTAAAGCCTACCACAGGAG
TATCAAGAGACGGGATTTCGAAGGAAATGGTATAATGAAGAACATGCGTGGTTCCAATGAA
GATCATCTTGGTGATCATTGCGACACACTTCGGTTTGATGATTTCACTTCAACAATTAAT
GAGAACCATCGTCATCATCAAGACCTGAAACAGAACATGACATTGCTGGAGAGTTATTAT
TCTTCTTTATCGTCCATCAATAGCGATTGGCCAGCTTGTCTTCTCCAGTACAACCTGA

>G1360 Amino Acid Sequence (conserved domain in AA coordinates: 18-174)

MGRNNDGDQKMEDVLLPGFRFHPHTDEELVSFYLRKRVQHNPLSLIELIRQLDIYKYDPWD
LPKFAMTGEKEWYFYCPRDRKYRNSRPNRVTGAGFWKATGDRPIYSSEGNKICIGLKKS
LVFYKGRAAKGVKTDWMMHEFRLPSLSESPSPSKRFFDSPVSPNDSWAICRIFKKTNTTT
LRALSHSFVSSLPETSTDMSNQKQSNTRYHFSDDKILKPSHFQFHENMNTPKTSNST
TPSVPTISPFYSLDFTSYDKPTNVFNPVSLDQOYLTLNLFATQETQPPRLPSSNEIP
SFLNLTSSDSTFLGEFTSHIDLSAVLAQEQCPPLVSLPQEQETGFEGNGIMKNMRGSNE
DHLGDHCDTLRFDDFTSTINENHRHHQDLKQNMTLLESYSSSLSSINSDLPACFSSTT*

>G1364 (1..537)

ATGGCGGAGTCGCAGGCCAAGAGTCCCGGAGGCTGTGGAAGCCATGAGAGTGGTGGAGAT
CAAAGTCCAGGTCGTTACATGTTCTGTGAGCAAGATAGGTTTCTTCCGATTGCTAACATA

AGCCGTATCATGAAAAGAGGTCTTCTCTGCTAATGGGAAAATCGCTAAAGATGCTAAGGAG
ATTGTGCAGGAATGTGTCTCTGAATTCATCAGTTTCGTCACCAGCGAAGCGAGTGATAAA
TGTCAAAGAGAGAAAAGGAAGACTATTAATGGAGATGATTGCTTTGGGCAATGGCTACT
TTAGGATTTGAAGACTACATGGAACCTCTCAAGGTTTACCTGATGAGATATAGAGAGGGT
GACACAAAGGGATCAGCAAAAAGGTGGGGATCCAAATGCAAAGAAAGATGGGCAATCAAGC
CAAAATGGCCAGTTCTCGCAGCTTGCTCACCAGGTCCTTATGGGAACTCTCAAGTAAC
TTTCTCTCTCTCTCTTACACTCAAGCAATACGCATCATTCTCTTCTAATTTGTTAA
>G1364 Amino Acid Sequence (conserved domain in AA coordinates: 29-120)
MAESQAKSPGGCGSHESGGDQSPRSLHVREQDRFLPIANISRIMKRGLPANGKIAKDAKE
IVQECVSEFISFVTSEASDKQREKRKTINGDDLLWAMATLGFEDYMEPLKVYLMRYREG
DTKGSAGKGD PNAKKDQSSQNGQFSQLAHQGPYGN SQVTFPLFSSHSNTHSLLIC*
>G1379 (68..622)
CTCTGCCTCTCTCTCTCTCAAAACCCATCTCGAAAGTCTTTCTCTTTTCGAGGGTTTAG
ATCCTCCATGGAAGGCGGCGGAGTTGCTGACGTGGCTGTCCCGGTACGAGGAAGAGAGA
CAGACCTTACAAGGAATTAGGATGAGGAAGTGGGAAAGTGGGTGGCGGAGATTCGTGA
GCCTAACAGCGCTCTAGGTTATGGCTTGCTCTTACTTACTCCCGAGCGCGCGCGCG
AGCTTACGACACGGCGGTTTCTATCTTAGAGGACCTACGGCGAGGCTTAACCTTCCCTGA
GCTTCTCTCTGGGGAGAAATCTCCGACGAGGATATGTCGGCTGCGACCATCAGGAAGAA
AGCCACGGAGGTCGGTGTCTCAGGTTGATGCTTTGGGCACGGCGGTGCAAAATAACGCCA
CCGTGTTTTTGGTCAGAAATCGAGATAGTGATGTGGATAATAAGAATTTTCATCGGAATTA
TCAAACGGTGAACGAGAAGAAGAAGAAGATGAGGATGACAAGAGATTGAGGAGTGG
CGGCCGTTATTGGATCGGGTTGACTTGAATAAATTACCCGACCCGAAAGCTCCGATGA
AGAATGGGAAAGCAAACATTAAAAATATATAGTTTGGAGCGGTGGCTGTGCTAACGTAC
GCCAACGGCTTGCTTCTACGAATCATTAGCGCCGTTTTATGATTTTTTTTTTTTTTTTT
CATTATCTGAAAATTTAGGGCTTTTTAGTTATTAATTTTTGTTTTGTTTTTCTTTCT
TGCGAGTTTTTGCGGTTTTATGGAATTTTAGGCTATTGCTTAACGAAAAAAAAAAAAAA
>G1379 Amino Acid Sequence (domain in AA coordinates: 18-85)
MEGGGVADVAVPGTRKDRPYKGI RMRKWKWVAEIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTARLNFPELLPGEKFSDEDM SAATIRKKATEVGAQVDALGTAVQNNRHRV
FGQNRSDVDNKNFHRNYQNGEREEEEDEDDKRLRSGRLLDRVDLNLKLPDPESDDEW
ESKH*
>G1384 (33..977)
GTACATTTTTTTTTTGTTATTTTCAGGAACTCCGATGGCGGATCTCTTCGGTGGTGGCCACG
GCGGCGAGCTTATGGAAGCACTTCAACCTTTTTACAAAAGTGCTTCCACGTCTGCTTCAA
ATCCTGCGTTTGCGTCTCTCAAACGATGCGTTTGCGTCTGCCCCAAACGACCTATTTTCTT
CTTCTCTTACTATAATCTCATGCATCTTTATTCCCTTACATTCCACAACCTCTTACC
CGGATATTTATTCTGGATCCATGACCTATCCATCTTCAATTCGGGTGCGATCTTCAACAAC
CCGAAAACCTACCAATCTCAGTTCCATTACCAAAACACTATCACTTACACTCACCAAGACA
ACAACACTTGCAATGCTTAACCTTATTGAGCCGAGCCAAACCGGTTTTATGACCCAAACCGG
GTCCGAGTTTCGGGTTTCGGTTTCAAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGGAAAATGGGTGCGCGGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTCGACACGGCTGAAGAAGCCGCTTGGCTTATGATCGCGCCGCTTAAAGCTTC
GTGGTGACTCGGCTCGGCTTAACTTCCAGCTCTCCGATACCAACCGGCTCGTCTCCGT
CTGATACCGGCGAATATGGTCTTATTCAAGCTGCCGTAGACGCTAAACTAGAAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCGCGGCTTCTTCACTGAGCAGCCGTCAGCGCCACAACAACATTCCGGGTGCGGTGAAA
GTGATGGGTGCGGTTTCCCGACTTCGGATGTTATGGTGCAGGAGATGTGCAAGAGCCAG
AGATGCCATGGAATGAAAATTTATGCTCGGCAAGTGTCTTCTTATGAGATGATTGGG
CTTCAATTTTATCGTGAAAAATTAGGATTCAATTCAATTTTATTCATTTTAACTTGTTTG
TATTTTCTTTTAACTTTAGGGTTATTAGCTGTGCGTAA
>G1384 Amino Acid Sequence (domain in AA coordinates: TBD)
MADLFGGGHGGELMEALQPFYKSASTSASNPAFASSNDAFASAPNDLFSSSSSYNPHASL
FP SHSTTSYPDIYSGSMTYPSSFGSDLQFPENYQSQPHYQNTITYTHQDNNTCMLNFIEP
SQPGFMTQPGSSGSVSKPAKLYRGVRRHWGKWVAEIRLPNRRLRLWLTGTFDTEEAAL
AYDRAAFKLRGDSARLNFALRYQTGSSPSDTGEYGP IQAAVDKLEAILAEPKNQPGKT
ERTSRKRAKAAASSAEQPSAPQOHSGSGESDGSPTS DVMVQEMCQEPMPWNENFMLG
KCPSYEIDWASILS*

>G1399 (261..1475)

AGGTCTGAATTTTCTGAAATTAAGATTTCATTCCTCCATGGAAGAAGCTCTGTTTTTATTCT
CTTTAGCTTAGCTTAGCTTCTACTGATCTGTTTTTGGCTACAAAATCCCATCTTTTCTTT
AAAACTCTTTATCTCTGAATCTTGAGTTTCTTGTAAGAAGAAGCAATTTTGAATCTTT
CGTAATCATAAAGATTTCGTGGAGGATCTCTACTGATTTGTCTGGAATCTCTACTACAGAA
TCACTTGATCTTATGTCCGGATGGAGGAGAGAGAAGGAACCAACATCAACAACAACATCA
CTAGCAGTTTTCGGCTTGAAGCAGCAACATGAAGCTGCTGCTTCTGATGGTGGTTACTCAA
TGGACCCACCACCAAGACCCGAAAACCCCTAACCCGTTTTTAGTCCCACCCACTACTGTCC
CCGCGGCCCGCCACCGTAGCAGCAGCTGTTACTGAGAATGCGGCTACTCCGTTTAGCTTAA
CAATGCCGACGGAGAACACTTCAGCTGAGCAGCTGAAAAAGAAGAGAGGTAGGCCGAGAA
AGTATAATCCCGATGGGACTCTTGTCGTGACTTTATCGCCGATGCCAATCTCGTCCTCTG
TTCCGTTGACGTCGGAGTTTCTTCCAAGGAAACGAGGAAGAGGACGTGGCAAGTCTAATC
GATGGCTCAAGAGTCTCAAATGTTCCAAATTCGATAGAAGTCTGTTGATACCAATTTGG
CAGGTGTAGGAACTGCTGATTTGTTGGTGCCAACTTTACACCTCATGTACTGATCGTCA
ACGCCGGAGAGGATGTGACGATGAAGATAATGACATTCTCTCAACAGGATCTCGTGCTA
TCTGCATCTTTTCACTAATGGTCCCATCTCCAATGTTACGCTTCGTCAATCTATGACAT
CCGGTGGTACTCTAAGCTTATGAGGGTCGTTTTGAGATTCTCTCTTTGACGGGTTCTGTTA
TGCAAAATGACTCTGGAGGAACCTCGAAGTAGAGCTGGTGGTATGAGTGTGCTTGCCTG
GACAGATGGTCTGTCTTTGGTGGAGGACTCGCTGGTCTCTTTCTGCTGCTGGTCTCTG
TCCAGGTAATGGTAGGGACTTTTATAGCTGGTCAAGAGCAGTCACAGCTGGAGCTAGCAA
AAGAAAGACGGCTAAGATTTGGGGCTCAACCATCTTCTATCTCTTAAACATATCCGCAG
AAGAACGGAAGCGGAGATTCCGAGAGGCTTAACAAGTCTGTTGCTATTCTGACCAACCA
CTTCATACACGCATGTAACACAACAATGCGGTTTACAGTTACTATACAACTCGGTTA
ACCATGTCAAGGATCCCTTCTCGTCTATCCAGTAGGAGGAGGAGGAGGTGGAGAGGTAG
GAGAAGAAGAGGCTGAAGAAGATGATGAATTAGAAGGTGAAGACGAAGAATTCCGAG
GCGATAGCCAACTCGACAACGAGATTCCGAGCTGATGATGATCATACGGTTTCTTTTCGC
GGATTTGTTAGTTTGTATGGATTTCAGATTTTGGTTGATTGTTTTTATTAACACAGAATG
TTTGAAGCTGCTATCTTTAGGTTCCCATCTCTTGTGATTGTTGAGTATCCTTGTAGA
AACAAACTTACTGTTGCAAACTCTCTTCAAAAAAGTTTCACTTTGCTTTCCCA

>G1399 Amino Acid Sequence (domain in AA coordinates: 86-93)
MEEREGTINNNITSSFGLKQHEAAASDGGYSMDPPRPNPNPFLVPPTTVPAAATVA
AAVTENAATPFSLTMPENTSAEQLKKRGRPRKYNPDGTLVVTLSMPPISSSVPLTSEF
PPRKRGRGRGKSNRWLKKSQMFQFDRSPVDNLAGVGTADFVGANFTPHVLIVNAGEDVT
MKIMTFSSQGSRAICILSANGPISNVTLRQSMSTGGTLTYEGRFEILSLTGSFMONDSGG
TRSRAGGMSVCLAGPDGRVFGGGLAGLFLAAGPVQVMVGTFIAGQEQSQLELAKERRLRF
GAQPSISIFNISAEERKARFERLNKSAIPAPTTSTYHVNNTTNAVHSYYTNSVNHVKDPF
SSIPVGGGGGEGVEEDDDLEGEDEEFGGDSQSDNEIPS*

>G1415 (60..680)

CCTTATCACTACCAAAAGTCGTACATAATATCACTTTTCGAGTTATCAACATCCGTACA
TGTCATCCATAGAGCCAAAAGTAATGATGGTTGGTGCTAATAAGAAACAACGAACCGTCC
AAGCTAGTTTCGAGGAAAGTTGTATGAGAGGAAAAGGTGGACCCGATAACGCGTCTTGCA
CTTACAAAGGTGTTAGACAACGCAC'TTGGGGCAAATGGGTGCTGAGATCCGCGAGCCTA
ACCGAGGAGCTCGTCTTTGGCTCGGTACCTTCGACACCTCCCGTGAAGCTGCCTTGGCTT
ATGACTCCGAGCTCGTAAGCTCTATGGGCTGAGGCTCATCTCAACCTCCCTGAGTCCT
TAAGAAGTTACCCTAAAACGGCGCTCGTCTCCGGCGTCCCAGACTACCAAGCAGCAACA
CCGGTGGAAAAGCAGCAGCGACTCTGAGTCGCCGTGTTTATCCAACGAGATGTCATCAT
GTGGAAGAGTGACAGAGGAGATATCATGGGAGCATATAAACGTGGATTGCGCGTAATGG
ATGATTCCTCAATATGGGAAGAAGCTACAATGTCGTTAGGATTTCCATGGGTTCATGAAG
GAGATAATGATATTTCTCGGTTTGATACTTGTATTTCCGGTGGCTATTCTAATTGGGATT
CCTTTTCACTCCCACTTTGAGGTGTCACTAGACTCTCTTAATTGTTAAGTTATCATATA
CAAACCTACATATATATACAAATATAGTCACCGTGAAGTATAGTAAATAAACA
CCAGTTACATGTACTTATATATGTGCACATCTATATATGTGGTTTGTCTGTATAGTGA
AAGCAGATTCTTACCATATCA

>G1415 Amino Acid Sequence (domain in AA coordinates: TBD)
MSSIEPKVMVMGANKKQRTVQASSRKGMRGKGGPDNASCTYKGVQRQTGWKVAEIREP
NRGARLWLGTFTDSREAAALAYDSAARKLYGPEAHLNLPESLRSYPKTASSPASQTTPSSN
TGGKSSSDSESPCSSNEMSSCGRVTEEISWEHINVDLPVMDSSIWEEATMSLGFPPVHE

GDNDISRFDTICISGGYSNWDSFHSP*
>G1417 (32..1501)

TCTATCTCTATCTATCTCTCTTTGTCTGCAATGGAAGAACATATTCAAGATCGCCGTGA
AATTGCGTTCTTACACTCAGGAGAATTTCTCCACGGAGATTCTGACTCAAAGGATCATCA
ACCGAACGAGTCTCCGGTGAACGTCATCACGAGTCGTCTATCAAAGAAGTTGATTTCTT
CGCTGCTAAAAGTCAGCCGTTTGATCTTGGTCATGTGAGAACAACGACGATCGTTGGATC
ATCTGGTTTAAATGATGGATTAGGTTTGGTAAATTCATGTCATGGAACATCAAGCAATGA
TGGCGATGACAAAACAAAACCTCAAATTAGTAGACTGAAGTTGGAGCTAGAGAGGCTTCA
CGAGGAGAATCACAACTGAAGCATTATTAGATGAGGTCAGTGAGAGTTACAACGACCT
CCAAAGAAGAGTTTTTGTAGCAAGACAAACACAAGTGAAGGCTTTCATCATAAAACA
TGAGGATGTACCTCAAGCTGGTTCTCACAAGCTCTAGAGAACAGAAGACCAAAGGATAT
GAACCATGAAACTCCGGCCACCACCTTGAACGACGGTCTCCAGACGACGTGGATGGTCG
TGATATGCACCGAGGATACCAAAAACCTCTCGAATAGACCAAAACAAGGACTAATCA
TGAAGAACAACAACCCCTCATGATCAATTACCCTATAGAAAAGCTAGGGTTTCCGTTAG
AGCTAGATCTGATGCCACTACGGTAAATGACGGATGTCAATGGAGAAAATACGGTCAGAA
AATGGCGAAAGGGAATCCATGTCTCGCGCTTATTATCGTTGCACCATGGCCGTTGGATG
TCCTGTCCGTAAACAGGTCAACGATGCGCGGAGGATACAACCTATCTTGACAACAACGTA
CGAAGGAAACCATAACCATCTCTTCCCCCGTCAGCCACAGCCATGGCTGCAACCACCTC
CGCCGAGCAGCCATGCTCTTATCAGGCTCTCTCCAGCAACCTCCACCAAAACACTCTC
TAGCCCTCCGCCACGTCATCATCTCTTCTACCATAACTTCCCATACACCTCCACAAT
CGCAACACTCTCTGCCTCAGCTCCTTTCCCCACCATAACCTTAGACCTACCAACCCACC
TCGACCGCTACAACCGCCACCGCAGTTTCTAAGCCAGTATGGTCCCGCCGCTTTTTACC
AAACGCTAATCAAATTAGGTCTATGAATAATAATAACCAGCAGTTATTAATACCTAATTT
GTTTGGCCCAACAGCCCAACGTCGAAATGGTCGATTGAGTTAGGGCTGCGATTGCGAT
GGATCCGAACCTTACGGCGGCACCTTGGCGCGCGATCTCAAACATTATCGGAGGAGGTAA
TAACGACAACAATAATAACTGATATTAATGATAACAAGGTTGATGCAAAAAGTGGAGG
GAGTAGTAACGGAGATTGCCACAGCTTCTCAGTCTTGCAACCTTTCTCTACAACTA
ATTTTACTACCATTATTATATGTTATCTTATTATATATTACACACATATTATACATTA
TGCCTATCTTAAGTTTTTTTTTGGGGGCCATTATATATGAATGATATGGAGATCACTGAG
AGAGAGAGAGAGCTATTATGGGTTTTTTTTT

>G1417 Amino Acid Sequence (domain in AA coordinates: 239-296)
MEEHIQDRREIAFLHSGEFLHGDSDSKDHQPNESPVERHHESSIKEVDFFAAKSQPFDLG
HVRTTTIVGSSGFNDGLGLVNSCHGTSSNDGDDKTKTQISRLKLELERLHEENHKLKHL
DEVSESYNDLQRRVLLARQTQVEGLHHKHQEDVPQAGSSQALENRRPKDMNHETPATTLK
RRSPDDVDGRDMHRGSPKTPRIDQNKSTNHEEQQNPDPQLPYRKARVSVRARS DATTVND
GCQWRKYGQKMAKNPCPRAYYRCMAVGCPVRKQVORCAEDTILTTTYEGNHNHPLPP
SATAMAATTSAAAAMLLSGSSSNLHQTLSPPSATSSSSFYHNFYPYSTIATLSASAPFP
TITLDLTNPPRLPQPPQFLSQYGPAAFLPNANQIRSMNNNNQQLLIPNLFPGPQAPPREM
VDSVRAAIAMPNFTAALAAAIISNIIGGGNNNNNTDINDNKVDAKSGSSNGDSPQLP
QSCCTTFSTN*

>G1442 (1..1293)

ATGGGAACAAGAGCAGAACGCAAGGAAGATTTTGTGGTGGGTTGGATTTGGTGTGTA
GAAAATTTCGCATAAAGACGTTATGGTGCTACCTCATCATCACTATTATCCATCATATTCA
TCACCTTCTCTTCTTCTTTGTGTTACTGTTCTGCTGGTGTAGCGATCCCATGTTCTCT
GTTTCTAGCAATCAGGCTTACACTTCTTCTCACAGTGGTATGTTACACCCGCGGTTCT
GGTCTGCTGCTGTGACTGTAGCAGATCCTTTTTTCTCCTTGAGCTCTTCAGGGGAAATG
AGAAGAAGTATGAACGAAGATGCTGGTGCAGCTTTCAGCGAAGCTCAATGGCATGAGCTT
GAGAGGCAGAGGAATATATACAAGTACATGATGGCTTCTGTTCTCTTCTCCAGAGCTT
CTCACACCTTTTCCCAAGAACCACCAATCAAACACTAACCCGGATGTAACGTGGCAGTG
GCGACAGGAGGCTCATTCAGCTGGGGATTGCTTCAAGCGCAAGCAATAACACGCTGAT
CTGGAGCCATGGAGGTGCAAGAGAACAGATGGGAAGAAATGGAGGTGCTCTAGAAACGTG
ATTCTGTGATCAGAAATACTGTGAGAGACACACACAAGAGCCGCTCCTCGTTCAAGAAAG
CATGTGGAATCATCTCACCAATCATCTCACCACAATGACATTCGTACGGCTAAGAATGAT
ACTAGCCAGCTTGTGAGAACTTATCTCAGTTTACGGACAACCTATAAGCCAGATCCCT
GTGCTTTCTACTCTTCCGTCTGCTCTCTCCATATGATCACCACAGAGGACTGAGGTGG
TTTACGAAAGAGATGATGCCATTGGAACCTTAAACCCGAGACTCAAGAAGCTGTCCAG
CTGAAAGTTGGATCAAGCAGAGAGCTCAAACGGGGATTGATTATGATCTGAATTTTCAGG

CAGAAAGAGCCAATAGTAGACCAGAGCTTTGGAGCATTGCAGGGTCTATTAAGTCTAAAC
CAGACACCACAACATAACCAAGAAACAAGACAGTTTGTGTAGAGGAAAGCAAGATGAA
GCGATGGGAAGCTCTCTGACACTCTCAATGGCTGGAGGAGGCATGGAGGAAACAGAGGGA
ACAAACCAGCATCAGTGGGTTAGCCATGAAGGTCCATCATGGCTCTATTCAACAACACCA
GGTGGACCATTGGCTGAAGCACTGTGTCTCGGTGTCTCCAACAACCCAAGTTCTAGTACT
ACTACTAGTAGCTGCAGCAGAAGCTCAAGCTAA

>G1442 Amino Acid Sequence (domain in AA coordinates: 172-223)
MGTRAERKEDFVGGFGFVGVVENS HKDVMVLP HHHYPSYSSPSSSLCYCSAGVSDPMFS
VSSNQAYTSSHS GMFTPAGSGSAAVTVADPFFSLSSSGEMRRSMNEDAGAAFSEAQWHEL
ERQRNIYKMMASVPVPPELLTPFPKNHQSNTPDVTVA VATGGSLQLG IASSASNTAD
LEPWRCRKT DGGKWRCSRNVIPDQKYCERH THKSRPRSRKHVESSHQSSHHNDIR TAKND
TSQLVRTYPQFYGPISQIPVLSTLPSASSPYDHRGLRWFTKEDDAIGTLNPETQEAVQ
LKVGSRELKRGFDYDLNFRQKEPIVDQSFGALQGLLSLNQTPQHNQETRQFVVEGKQDE
AMGSSLTSLMAGGMEETEGTNQHQWVSHEGPSWLYSTTPGGPLAEALCLGVSNPNSSST
TTSSCSRSSS*

>G1454 (86..1180)
CTAGTAGTGATGATATGATCGCTTCTTCTCTCTACAATCTCAGAAACCTCCGATCACGGTT
TTAGATATCTTCTACAACGGATAACAATGGAGAGACCGATTCTTCCGGTGGTCCACCACC
GCCACAACCTAACCTTCTCCAGGCTTCCGGTTTACCCTACCGACGAAGAGCTTGTGT
TCACTACCTCAACGCAAGCAGCCTCTGCTCCTTTACCTGTGCGCATCATCGCCGAAGT
CGATCTCTATAAATTTGATCCATGGGAACCTCCCGCTAAAGCATCGTTTGGAGAACAAGA
ATGGTACTTCTTTAGTCCACGAGATCGGAAGTATCCAACGGAGCAAGACCAACACAGAGC
GGCGACTTCAGGTTATTGGAAAGCGACCGGTACAGATAAACCGGTACTTGTCTCCGACGG
TAACCAAAAGGTGGCGGTGAAGAAGGCACTAGTCTTCTACAGTGGTAAACCAACCAAAAGG
CGTTAAAGTGATTGGATCATGCATGAGTATCGTCTCATCGAAAACAAACCAACAATCG
ACCTCCTGGCTGTGATTTCCGCAACAAAAAACTCACTCAGACTTGATGATTGGGTGTT
ATGTAGAATCTACAAGAAGAACACGCAAGTCGACATGTTGATAACGATAAGGATCATGA
TATGATCGATTACATTTTCAGGAAGATTCCTCCGTCTTTATCAATGGCGGTGCTTCTAC
AGGACTTCACCAACATCATCATAATGTCTCAAGATCAATGAATTTCTTCCCTGGCAAATT
CTCCGGTGGTGGTTACGGGATTTTCTCTGACGGTGGTAACAGAGTATATACGACGGCGG
TGCGATGATCAACAATAATTGGTACTGACTCAGTAGATCACGACAATAACGCTGACGTCGT
TGGTTTAAATCATGCTTCGTCGTCAGGTCCTATGATGATGGCGAATTTGAAACGAATCT
CCCGGTGCCGTATTGGCCTGTAGCAGATGAGGAGCAAGATGCATCTCCGAGCAAACGGTT
TCACGGTGTAGAGGAGGAGGAGGAGATTGTTGCAACATGCTTCTCCATGATGGAAGA
GACTCCACCATTGATGCAACAACAAGGTGGTGTGTTAGGAGATGGATTATTGAAACGAC
ATCGTACCAATTACCCGGTTTAAATTGGTACTCTTCTTAATCAAATGTGTTTCGCCGCCG
GTGTGAAGAAATTTTCGGTGACAGTGAAGATTTTTTTCGATTGGTGGGGTCATTTGCAT
GCATTATATAAATTTGAGATTTGTGTATATGTTTGGGTTAATTAATTGTCACAGGGGC
>G1454 Amino Acid Sequence (conserved domain in AA coordinates: 9-178)
MESTDSSGGPPPPQPNLPPGFRFHP TDEELVVHYLKRKAASAPLPVAIIAEVDLYKFDPW
ELPAKASFGEQEWYFFSPRDRKYPNGARPNRAATSGYWKATGTDKPVLASDGNQKVG VKK
ALVFYSGKPPKGVKSDWIMHEYRLIENKPNRPPGCD FGNKKNLRLDDWVLCRIYKKN
ASRHVDNDKDHDIMIDYIFRKIPPSLSMAAAS TGLHQHHNVSRSMNFFPGKFSGGGYGIF
SDGNTSIYDGGMINNIGTDSVDHDNNADVGLNHASSSGPMMANLKR TLPPVYPVPA
DEEQDASPSKRFHGVGGGGDCSNMSSSMMEETPPLMQQGGV LGDGLFR TTSYQLPGLN
WYSS*

>G1459 (1..1272)
ATGATGAAAGGTCTGATTGGGTATAGATTTAGTCCGACGGGAGAGGAAGTGATCAACCAT
TACCTAAAGAACAACCTTCTGGGTAAAGTATTGGCTCGTTGATGAAGCTATTAGCGAGATC
AACATCTTGAGTCACAAACCCAGCAAGGATTGCGCTAAGTTAGCTAGGATCCAATCGGAA
GATCTTGAATGATATTCTTCTCTCCGATTGAGTACACGACCCGAATAAGATGAAAATG
AAGAGGACGACAGGTTCTGGGTTTGGAAACCTACTGGTGTGATCGGGAAATTAGGGAT
AAAAGAGGAAATGGTGTGTGATAGGGATTAAAGAAGACGCTTGTGTACCATGAAGGTAAG
AGTCCTCATGGAGTTAGAACTCTTGGGTTATGCACGAGTATCACATCACTTGCTTGCTT
CATCATAAGAGGAAATATGTTGTCTGCCAAGTAAAGTATAAGGGTGAAGCTGCAGAAATT
TCATATGAGCCAAGTCCCTCTTGGTATCCGATTTCGCATACCGTCATAGCGATTACCGGA
GAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAGGTAAAGAAAATCTCTGGGTATGTCT

GTAGATGATTTGATAGAACCAATGAACCAACAAGAGGAGCCACAAGGTCCTCACTTAGCT
CCGAATGATGATGAGTTTATACGTGGATTGAGGCATGTTGATCGAGGGACGGTTGAATAT
TTGTTTGCCAATGAAGAAAACATGGATGGTTTGTCTATGAATGACTTGAGAATCCCAATG
ATCGTCCAACAAGAGGATCTCTCTGAGTGGGAGGGATTTAACGCAGACACCTTTTTCAGC
GACAAACAATAAAGTATAACCTTAACGTGCATCATCAACTAACGCCTTACGGCGATGGC
TATTTGAATGCATTTTCGGGTATAACGAAGGGAATCCTCCCGATCACGAATTAGTGATG
CAAGAGAACCGCAACGATCACATGCCAAGGAAACCTGTGACAGGGACCATTGATTATAGC
AGCGATAGTGGCAGTGATGCTGGATCCATATCTACAACGGTGAAACAAGAAATCCCAAGA
GCTGTTGATGCACCCATGAACAATGAGTCATCTTTGGTGAAACAGAGAAGAAAGGCTTG
TTTATTGTAGAGGACGCAATGGAGAGAAACCGCAAGAAACCACGATTTATCTATCTCATG
AAGATGATCATAGGCAACATCATATCGGTTTTACTACCCGTCAAAAGATTGATCCCGGTG
AAGAAGTTATGA

>G1459 Amino Acid Sequence (conserved domain in AA coordinates:10-152)

MMKGLIGYRFSPTGEEVINHYLKNKLLGKYWLVD E A I S E I N I L S H K P S K D L P K L A R I Q S E
D L E W Y F F S P I E Y T N P N K M K M K R T T G S G F W K P T G V D R E I R D K R G N G V V I G I K K T L V Y H E G K
S P H G V R T P W M H E Y H I T C L P H H K R K Y V V C Q V K Y K G E A A E I S Y E P S P S L V S D S H T V I A I T G
E P E P E L Q V E Q P G K E N L L G M S V D D L I E P M N Q Q E P Q G P H L A P N D D E F I R G L R H V D R G T V E Y
L F A N E E N M D G L S M N D L R I P M I V Q Q E D L S E W E G F N A D T F F S D N N N N Y N L N V H H Q L T P Y G D G
Y L N A F S G Y N E G N P P D H E L V M Q E N R N D H M P R K P V T G T I D Y S S D S G S D A G S I S T T V K Q E I P R
A V D A P M N N E S S L V K T E K K G L F I V E D A M E R N R K P R F I Y L M K M I I G N I I S V L L P V K R L I P V
K K L *

>G1460 (87..995)

CGTCGACCTTCACTCAAACCTAATCCCGGGAACCCGGAATTTTGATCATTTTGTCTTCT
TTTCGATCTGTTTCTATTTTAAAAAGATGATGAAAGATCCGACTGGGTATAGATTTAGTC
CGACGGGAGAGGAAGTGATAAACCATACCTAAAGAACAAAATTCTGGGTAAAGACTTGGC
TCGTTGATGAAGCCATTAGCGAGATCAACATCTTGAATCACAAACCCAGCAAGGATTTGC
CTAAGTTAGCTAGGATCCAATCGGAAGATCTTGAGTGGTACTTTTTCTCTCCGATTGAGT
ACACGAACCCGAATAAGATGAAATGAAGAGGACGACAGGTTCTGGGTTTTGGAAACCTA
GTGGTGTGATCGGAAAATTAGGGATAAAAGAGGAAATGGTGTGTGATAGGGATTAAGA
AGACGCTTGTGTACCATGAAGGTAAGAGTCCCTCATGGAGTTAGAACTCCTTGGGTTATGC
ACGAGTATCATCATCTTGCTTGCCCTCATATAAGAGGAAATATGTTGTCTGCCAAGTAA
AGTATAAGGGTGAAGCTGCAGAAATTTATATGAGCCAAGTCCCTCTTTGGTATCCGATT
CGCATACCGTCATAGCGATTAAACGGAGAACCGGAACCTGAGCTTCAGGTTGAGCAGCCAG
GTAAAGAAAATCTCTTGGGTATGTCTGTAGATGATTTGATAGAACCAATGAACCAACAAG
AGGAGCCACAAGGTCCTCACTTAGCTCCGAATGATGATGAGTTTATACGTGGATTGAGAC
ATGTTGATCGAGAGCCGTTGAATATTTGTTTGCCAATGAAGAAAACATGGATGGTTTGT
CTATTATGAATGACTTGACAATCCCAATGATCGCCCAACAAGAGGATCTCATTCTCTCTG
AGTGGGAGGGATTTATCGCAGCCACCTTTTTTCAGCGACAACAACAATAACAATAACCTTA
ACGTGCATCAACTAAGTCTTTCTTACCGGGATGATTATCAGAATGCATTTTGGGTTACA
ACGGAGCGNCCGCT

>G1460 Amino Acid Sequence (domain in AA coordinates: TBD)

MMKDPTGYRFSPTGEEVINHYLKNKILGKTWLVD E A I S E I N I L N H K P S K D L P K L A R I Q S E
D L E W Y F F S P I E Y T N P N K M K M K R T T G S G F W K P S G V D R K I R D K R G N G V V I G I K K T L V Y H E G K
S P H G V R T P W M H E Y H I T C L P H H K R K Y V V C Q V K Y K G E A A E I S Y E P S P S L V S D S H T V I A I N G
E P E P E L Q V E Q P G K E N L L G M S V D D L I E P M N Q Q E P Q G P H L A P N D D E F I R G L R H V D R E P V E Y
L F A N E E N M D G L S I M N D L T I P M I A Q Q E D L I L S E W E G F I A A T F F S D N N N N N L N V H Q L T S F L
P G *

>G147 (37..672)

AAATCATCAGATAGAAGGAAATATTCTGATTGAGAGATGGCTCGTGGAAGATTGAGCTT
AAGAGGATTGAGAACCCGGTTCACAGACAAGTGACTTTTTGCAAGAGGAGAACTGGTCTT
CTCAAGAAGGCTAAGGAGCTCTCTGTGCTCTGTGATGCCGAGATCGGTGTTGTGATCTTC
TCTCCTCAGGGCAAGCTCTTTGAGCTCGCTACTAAAGGAACAATGGAGGGAATGATTGAT
AAGTACATGAAGTGTACTGGTGGTGGTCTGTTCTTCTGCTACTTTTACTGCTCAA
GAACAACTTCAACCACCAATCTTGATCCGAAAGATGAGATCAACGTGCTTAAGCAAGAG
ATTGAGATGCTTCAGAAAGGGATAAGCTATATGTTTGGAGGAGGAGATGGGGCTATGAAT
CTTGAAGAACTTCTTTTGGCTTGAGAAGCATCTTGAGTATTGGATTTCTCAGATTCGCTCT
GCTAAGATGGATGTTATGCTTCAAGAAATTCAGTCATTGAGGAACAAGGAAGGAGTCCTC

AAAAACACCAACAAGTATCTCCTCGACAAGATAGAGGAAAACAACAATAGCATATTAGAT
GCTAACTTCGCAGTCATGGAGACAACTATTCCTATCCGCTAACAATGCCAAGTGAAATA
TTTCAGTTCTAGACCATAGGGTATTTGAAGACTATGTCTCACGAATTTAAATAACCTTGG
TAAGTATAATATAGTGTGTAAATCACACATAATTAATAAAGCCTGTGGAACCTCGC
TAGGCAGTTGAAAATCTATCCGTATGTTTTATCCTCTTGTGTTTACATTTGTTGGTGTGAA
GATGAAATGACTGCAAGTGTGGTGTGTACTTATAACTCTTCTACTTTCTATCTATGTTT
TGAATTTATGGATT

>G147 Amino Acid Sequence (domain in AA coordinates: 2-57)
MARGKIQLKRIENPVHRQVTFCKRRTGLLKKAKELSVLCDAEIGVVIFSPQGLFELATK
GTMEGMIDKYMKTCTGGGRGSSSATFTAQEQLPNLDPKDEINVLKQEIEMLQKGISYMF
GGGDGAMNLEELLLLEKHLEYWISQIRSAKMDVMLQEIQSLRNKEGVLKNTNKYLLDKIE
ENNNISILDANFAVMEITNYSYPLTMPSEIFQF*

>G1471 (1..735)
ATGGAGAACCAATCTATGTCTTCATCAAGCTCCTCCACACACAAACATGATCAAAAACCTC
AAAAGTTCCGTTGTGGCCATGGAGGTCCTGGAGGAGAAGGAGACAGTGAACAATCCGCCC
CAGTATTATAATAAGATCTACATCTGTTACTTGTGCAAGAGAGCGTTCCCAACCCCTCAT
GCCCTTGGCGGTACGGAACCAACCCACAAGGAGGACCGAGAATTGGAGAGGCAACAGATC
GAGTCAAGGCTTTCTAACAAGACAAGTCTAACTTGCTCTTGGTGGGTCTTCACAAGAT
GTTTTATCAAATGATAATCACCTTGGACTCTCTCTTGGTCCATTGAAGTCCATAGAAGGT
AGCAGCAGCAGCAACAACGTTAACCCTTGGCTTAATGTTGGAGTCCCTAGAGGAACCACA
GATATGAACATGAACAACCTATAGCTCACATGCTTTATCAACTGATGATATTAATCTTGAT
CTTACTCTTGGTCCATCTAAGTCCATAGGAGATAGCAACAATATCATTAAATAACAACACT
AACTCATCCTTCGATGGGAATCTGATCATTCCCGTTCGTCCTCGTGTGTCTAGATACCAT
TTTGTGCTGGGAACCCCTTGATTCAATCTCTAGAAACATTCCCTCCTTCTATTACTTTT
CCTCATCTAAACATCAATCTTTCTCATGATTCGTTTTCTTTACAAGAGAATGGTTCGGGC
TCTAGTCACTCATAA

>G1471 Amino Acid Sequence (domain in AA coordinates: 49-70)
MENQSMSSSSSTHKHDQKLKSSVAMEVLEEKETVNNPPQYNNKIYICYLCKRAPFPPH
ALGGHGTTHKEDRELERQOIESRLSNKDKSNLLFGGSSQDVLSDNHLGLSLGPLKSI
SSSSNNVNPLLNVGVPRTTDMNMNYSHALSTDDINLDLTLGPSKSI
GDSNNIINNNT
NSSFDGNLIIPVRPRVSRVYHFVAGNPLDSISRNPPTSITFPHLNINLSHDSFSLQENGSG
SSHS*

>G1475 (1..645)
ATGAAGAGAACACATTTGGCAAGTTTGTAGTAACAGAGACAAAACCAAGAAGAAGAAGGA
GAAGACGGTAATGGTGACAACAGAGTCATCATGAATCACTACAAGAATTACGAAGCTGGG
CTGATCCCATGGCCTCCCAAGAATTACACTTGCAGCTTCTGCAGGAGAGAGTTCAGATCT
GCTCAAGCACTTGGAGGCCACATGAATGTTCATAGAAGAGACAGAGCAAAACTCAGGCAG
ATCCCTTCTTGGCTCTTCGAACCTCACCACCACACACCTATTGCAAACCTTAACCTAAT
TTTAGCTCTTCTTCTTCTTCTTCAACAACAACAGCTCATCTTGAGCCTTCCCTAACCAAC
CAGAGATCCAAAACCACTCCTTTCTCTTCTGCCCCGTTTGATCTTTTGGACAGTACTACT
AGCTATGGAGGTTTGATGATGGACAGAGAGAAGAACAAGAGCAATGTATGTAGCAGAGAG
ATCAAGAAAAGTGCCATCGATGCATGTCATTAGTAAGATGTGAGATAAGCCGTGGGGAT
CTGATGAATAAGAAAGATGATCAAGTCATGGGGTTGGAGCTTGGGATGAGTTTGAGGAAT
CCCAACCAAGTTCTTGATTGGAGCTTCGACTAGGCTACCTCTAA

>G1475 Amino Acid Sequence (domain in AA coordinates: 51-73)
MKRTHLASFSNRDKTQEEGEDGNGDNRVIMNHYKNYEAGLIPWPPKNYTCSEFCRREFRS
AQALGGHMNVHRRDRAKLRQIPSWLFEPHHHTPIANPNPNFSSSSSSSTTAHLEPSLTN
QRSKTTFPFSARFDELDSSTSYGGLMMDREKNKSNVCSREIKKSAIDACHSVRCEISRGD
LMNKKDDQVMGLELGMSLRNPQVLDLELRLGYL*

>G1477 (1..606)
ATGTTGTCTCTCGACTCGAATTACGCTAGTGATATTAGCGACGATGCCTCCGCCACCGGA
TCGATAGAGAATCCTATATACAAATGCAAGTATTGTCTAGGAAGTTCGATAAAACACAA
GCATTAGGTGGTCATCAAAATGCACACAGAAAGGAGAGAGGTGAAAAACAACAAAAA
GCATTTTGGCGCATTTGAACCGACCAAGACAGATCTTTACGCGTACTCGTATTTCGTAT
CATCATTCATTTCTAACCAATACGCACTCCACCCGGGATTGTAACAGCCTCAGTACAAA
GTTGATAGATCATACAAGATGTCCATGGTCTACAACCAATATGTGGGATCCTCAAGCTCT
AGCTTTGCAGGACTACAAAGTGACCCAAGTCAAGGAATGAACCAGGATTGGACCTTTACC

GGGATCCCATTCTACCCCAATCTCAACCTCAACCCTATCGTCACCAATATGTTTGGAT
CTTTGCCTTGGCATTGGTAGCTCCCAACCCCAACCACAACCTCAAGAACCAAAATGATGCA
ACAGAAGAGATGGATGCTGAGAAAGAAAATGATGGTTCTTCCCTTTCTCTCTCACTCAAA
CTGTGA

>G1477 Amino Acid Sequence (domain in AA coordinates: 29-48)
MLSSDSNYASDISDDASATGSIENPIYKCYCPKFDKTQALGGHQNAHRKEREVEKQOK
AFLAHLNRPEPDLAYSYSYHHSFPNQYALPPGFEPQYKVDRSYKMSMVYNQYVGSSSS
SFAGLQSDPSQGMNQDWTFTGIPFLPQSQPQLSSPICLDLCLGIGSSQTQPQPQEPNDA
TEEMDAEKENDGSSLSLSLKL*

>G1487 (1..1020)

ATGGAACAAGCCCGCTTGAAGAGCAGCGTCAGGAAAGAGATGGCTCTCAAAACGACTTCT
CCGGTTTACGAAGAGTTTCTTGCCGTCACCACCGCTCAAAATGGCTTTTCCGTCGACGAT
TTCTCTGTAGACGACTTGCTTGACTTGTCAAACGATGACGTTTGTGCCGACGAAGAACT
GACCTCAAGGCTCAACATGAGATGGTCCGTGTTTCTCTGAGGAACCAACGACGACGGA
GACGCTCTTCGCCGGAGCAGCGATTTCTCCGGCTGTGACGACTTTGGTTCTCTCCCTACA
AGCGAACTCTCTCTTCGGCGGATGATTTAGCGAACCTTGAGTGGCTCTCTCATTTCTGTG
GAGGACTCCTTCACGGAATATTCGGGTCCAAACCTCACCGGAACCCGACTGAGAAACCG
GCGTGGTTAACGGGTGACCGGAAACATCTGTGACTGCAGTCACGGAAGAGACCTGTTTC
AAATCCCCGTGTTCGGGCTAAAGCCCGTAGCAAACGTAACCGCAATGGCCTCAAGGTCTGG
TCGCTTGGTTTCGTCTCTCTCGGGTCTCTCTCGTCCGGTTCGACCTCCTCCTCCTCT
TCGGGTCTTCCAGCCCGTGGTTCTCCGGCGCTGAGCTGCTCGAGCCTGTGGTCACGTCA
GAGAGGCCACCGTTTCCCAAGAAGCATAAGAAAAGTCAAGCCGAGTCTGTTTCTCCGGT
GAGCTGCAGCAGCTGCAACCTCAGCGAAAGTGACGCACTGCGGCGTTTCAGAAAACCTCCG
CAGTGGAGAGCCGGCCAATGGGAGCCAAGACCCTGTGCAATGCGTGCAGGTGTCCGGTAC
AAGTCCGGTAGGTTGCTACCGGAATACAGACCCGCTTGAGTCCCGACATTCTCGAGTGAG
CTGCACTCGAACCACCGGAAAGTCATAGAGATGAGGCGGAAGAAGGAGCCAACAGT
GACAACGAAACCGGTTTAAACCAGCTGGTTCAGTCCCCACAAGCTGTACCAAGTTTTTGA
>G1487 Amino Acid Sequence (domain in AA coordinates: 251-276).
MEQAALKSSVRKEMALKTTSPVYEEFLAVTTAQNFGSVDDFVSDDLDSNDDVFADDEET
DLKAQHEMVRVSSEEPNDDGDALRRSSDFSGCDDFGSLPTSELSLPADDLANLEWLSHFV
EDSFTEYSGPNLTGTPTEKPAWLTGDRKHPVTAVTEETCFKSPVPAKARSKRNRNGLKVV
SLGSSSSSGPSSSGSTSSSSSGPSPWFSGAELLEPPVTSERPPFPKHKHRSAAESVFSG
ELQQLQPQRKCSHCGVQKTPQWRAGPMGAKTLCNACGVRYKSGRLLPEYRPACSPTFSSE
LHSNHRKVIEMRRKKEPTSDNETGLNQLVQSPQAVPSF*

>G1492 (149..919)

AATCCCAACCCACACACCTCTCAAATCCTCCTCTCCTCGTTTCTCTTTCTCTCCTCTTCA
CAGAACCAAAACATATCAAACCTTTTTTCTCTTGGGTTTAAAGTAAAAATCGAATCTTTG
TGTCGGTTTTTAGGGTTCTTGAAACGATATGGGTAAGTCTAGTGGTAGAAATGGTAACGG
AAGCTTTAACGGCAATAAATTTACGGAGTTAGACCTTACGTACGGTCTCCAGTTCCACG
GCTTAGATGACGACCGGATCTTCAACGTTGTTTCTGTTACGCCGTCGAGATTCTCGGTGG
TCAACACCGAGCAACACCAAACTTGTCTTAAGATGATGGATGTGAAGGGACTTACCAT
TTCACATGTCAAAGCCACCTTCAGATGTATAGAGGAGGTTCAAAGCTCACTTTGGAGAA
ACCAGAAGAAAGCTCATCATCTTCAATAAGAAGAAGACAAGACAGTGAAGAAGATTATTA
TCTTCATGACAACTTGTCTTTACACACAAGGAATGATTGTCTTTTGGGTTTTCACTCTTT
TCCTCTTTCTTACATTCTTCATTTAGAGGAGGAGGAGGAGGAAGAACAAGAGCAGCA
GACTTCAGAGTCTGGTGGTTATGATGATGATGCTGACTTCTTACATCAAGAAGATGAA
CGATACGACGACGTTTTTGTTCACATCATTTCCCAAGGGAACAGAGGAGTGGCGGGAACA
AGAACACGAAGAAGAAGAAGATTTGTCTGTTGCTCTGTCTGTTAAATCATCATATTG
GAGAAGCAATGGATCAGCGTGGTGAGCGAAACGAGTGAAGCAGCAGTCTCGACTTGTTC
AGCACCATTCTGATCCAAAGATTGCTTTGGTTCTTCAAAGATTGATCTTAATCTGTCAAT
TTCTCTCTCGGTAGCTAAATAAGTTATGCAAGATTTAGGTTTCAGAGAACTATTTCGGAT
GTGTTTTTGAAGTATGATGTTAGTAGAGAAACCTAGAAAATGAAGTTTAGAT
AAATTATCAACGCAGCGTTTTGATCGCCTTTGAACGGAATAAATAA

>G1492 Amino Acid Sequence (domain in AA coordinates: 34-83)
MGKSSGRNGSGFNGNKFHGVPRPVPRPRLRWTPDLHRCFVHAVEILGGQHRATPKLV
LKMDVKGLTISHVKSHLQMYRGGSKLTLEKPESSSSSIRRRQDSEEDYYLHDNLSLHT
RNDCLLGFHSFPLSSHSSFRGGGGRTKEQQTSESGGYDDADFLHIKKMNDTTTFLSHH

TGTGGAGGCCAACTGCAATTGGCGAAATGACCTTCGAAGAGCACCATCTTCGCATCCTC
AACGCTCGTTTTGACTGAAGAGATCAAGCAACTTTCCGTGACAGCGGAAAAGATATCAAGG
CTTACGGGGATACCAGTAAGGAGCCATCCCCGTGTGTCTCTCTAATCCTCCTCCAAAT
TTCGAGTTCGGGATGGGATCTAAGGAAAATGTCGGAACCCTCGAGGGAAACCCTGGA
CCTGCAGATGCTAATAACCAAGCCGATCATCATGGAGTTGGCATTGGAGCCATGGAGGAG
CTCTTGGTGTAGGCTCAAGTGGCTGAACCCTGTGGATGGGAGGATTAAATGGCACTAGC
TTAGCTTTGAACTTGGATGAATACGAAAAGACGTTTCGCACGGGTCTCGGTCTAGACTT
GGCGGGTTTCGAACCGAGGCATCCAGGGAACTGCACTCGTGGCAATGTGTCTACTGGC
ATTGTTGAAATGCTCATGCAAGAGAATCTGTGGTCAACAATGTTTGGCGGAATTGTTGGT
AGAGCCAGGACTCATGAACAGATAATGGCTGATGCTGTGGAACTTCAATGGAAATCTC
CAAATAATGAGTGTGAGTACCAAGTGTCTTCCCCGCTAGTCACAACCCGCGAAAGCTAC
TTCGTCGCTACTGTAAAGCAACAAGGAGAGGGTTTGTGGGCGGTGGTTCGATATTTCCATC
GACCATCTCTCCCAAACATCAACCTAAAAATGTCGCGCGGACCCTCTGGATGTCTGATT
CAAGAAATGCATAGTGGTTACTCCAAGGTTACATGGGTGGAACATGTGGAAGTAGATGAT
GCAGGAAGTTACAGCATCTTTGAGAAATTAATCTGTACTGGTCAAGCTTTTGCTGCTAAC
CGCTGGGTTGGTACATTGGTACGCCAGTGTGAGCGGATATCTAGCATCTTGTGACAGAT
TTTCAATCTGTGATTCCGGTGTATCACATAACGCTAACTAACCATGGAAAGATGAGCATG
CTGAAGATAGCTGAGCGGATTGCGAGAACCTTCTTTGCTGGAATGACCAATGCGACGGGG
TCTACAATATTTTCTGGTGTGAAGGAGAAGATATCAGAGTGTGACAATGAAGAGCGTG
AATGATCCAGGAAAGCCTCCCGGTGTCATTATTTGTGACGCCACTTCTTTTGGCTTCTCT
GCTCCTCTAACACTGTCTTTGACTTCCTCAGAGAGGCTACTACCGACACAATTGGGAT
GTTCTCTGCAACGAGAGATGATGCACAAGATAGCAGAGATTACGAATGGGATAGACAAA
AGGAAGTGTGCAAGTTTACTCCGGCATGGACACACTAGCAAGAGCAAGATGATGATAGTT
CAAGAGACTTCTACTGACCAACAGCTTCATTTGTGCTTTATGCGCCTGTTGATATGACA
TCAATGGATATTACTCTCCATGGAGGTGGTGTATCTGACTTTGTGGTGATCCTGCCTTCT
GGTTTGTCTATTTTCCAGATGGTACGGGTAAGCCTGGAGGAAAAGAGGAGGATCACTT
TTGACCAATTCTCCTCCAAATGTCTGGTTGAGTCAAGTCTGAGGCTAGGCTGAGTGTAGC
TCTGTTGCAACTACTGAGAATCTGATTCTGTACAACCGTGGGAGGATCAAAGATTTGTTT
CCTTGTCAGACTGCTTGA

>G1544 Amino Acid Sequence (domain in AA coordinates: 64-124)

MSQSNMVPVANNGDNNNDNENNNNNNNNGGTDNTNAGNDSGDQDFDSGNTSSGNHGEGLG
NNQAPRHKKKKYNRHTQLQISEMEAFFREC PHPDDKORYDLQAQLGLDPVQIKFWFQNKR
TONKNQOERFENSELRNLNNHLRSENQRLREAIHQALCPKCGGQTAIGEMTFEEHHLRIL
NARLT EIKQLSVTA EKISRLTGI PVRSHPRVSPNPPNFEFGMSGKGNVGNHSRETTG
PADANTKPI IMELAFGAMEELLVMAQVAEPLWMGGFNGTSLALNLD EYEKTFRTGLGPRL
GGFRTEASRETALVAMCPTGIVEMLMQENLWSTMFAGIVGRARTHEQIMADAAGNFNGNL
QIMSAEYQVLSPLVTTRESYFVRYCKQQEGELWAVVDISIDHLLPNINLKR RRPSPGCLI
QEMHSGYSKVTWVEHVEVD DAGSYIFEKLI CTGQAFANRWVGT LVRQCERISSILSTD
FQSVDSGDHITLTHNGKMSMLKIAERIARTFFAGMTNATGSTIFSGVEGEDIRVMTMKSV
NDP GKPPGVII CAATSFWL PAPPNTVFD FLREATHRHNDVLCNGEMMHKIAEITNGIDK
RNCASLLRHGHTSKSKMMIVQETSTDPTASFVLYAPVDMTSM DITLHGGGDPDFVVLPS
GFAIFPDGTGKPGGKEGSLTISFQMLVESGPEARLSVSSVATTENLIRTTVRRIKDLF
PCQTA*

>G156 (39..755)

AGGAAGAGGGAGCCACTCATAAGAGGAAGAAGAGAGAGATGGGTAGAGGGAAGATAGAGA
TAAAGAAGATAGAGAATCAGACGGCGAGGCAAGTGACCTTCTCCAAGAGAAGAACTGGTC
TTATAAAGAAGACTCGTGAGCTCTCTATTCTCTGTGACGCTCACATCGGTCTCATCGTCT
TCTCAGCCACCGGAAAGCTTTCCGAGTTCTGCTCCGAACAGAACAGGATGCCTCAACTCA
TTGACCGATACTTGACATACCAACGGATTGCGACTTCTGATCATGACGACCAGGAGC
AATTGCACCATGAGATGGAACCTACTAAGAAGAGAGACATGTAACCTTGAGCTTCGTCTGC
GTCCATTCCATGGACATGACTTAGCCTCATTCTCTTAATGAGCTTGACGGACTCGAGA
GACAGCTAGAACATTCTGTCTCAAAGTCCGTGAGCGTAAGAGGAGGATGCTAGAAGAAG
ATAACAACAACATGTACCGTTGGCTTCATGAGCATCGTGCAGCGATGGAGTTTCAACAAG
CTGGGATAGATACCAAAACAGGGAGTATCAACAGTTTATAGAGCAGCTTCAGTGCTATA
AACCAGGGGAGTATCAGCAGTTTCTAGAGCAGCAGCAACAACAACCAACAGCGTTCTTC
AGCTTGCTACACTTCTTCTGAGATTGATCCTACTTACAATCTCCAGCTTGCTCAGCCTA
ATCTTCAAAACGATCCAACGGCCAGAATGATTAATACAATTCTCAATAGATATCTACTC

84

TGTTACTACTACGAGTTCATGCCTCTGAAGAACTGA
>G1587 Amino Acid Sequence (conserved domain in AA coordinates: 61-121)
MGYISNNNLINYLPLSTTQPPLLLTHCDINGNDHHQLITASSGEHDIDERKNNIPAAATL
RWNPTPEQITTTLEELYRSGTRPTTEQIQIASKLRKYGRIEGKNVFWFQNHKARERLK
RRRREGGAIKPHKDVKDSSSGHRVDQTKLCPSPHTNRPQPQHELDPASYNKDNANN
EDHGTTEESDQRASEVGKYATWRNLVWTSITQQPEEINIDENVNGEEETRDNRTLNLFP
VREYQEKTRGLIEKTKACNYCYFYMPLKN*
>G1588 (1..2232)
ATGTACCATCCAAACATGTTTGTAGAGCCATCATATGTTTCGATATGACCCCAAAGAGTACC
TCTGATAACGACTTGGGAATCACCGGTAGCCGAGAAGATGACTTTGAGACCAAGTCAGGT
ACCGAAGTCACTACTGAGAATCCTTCTGGTGAAGAGCTTCAAGATCCTAGCCAACGTCCC
AACAAAAAGAAGCGTTACCATCGCCACACGCAACGCCAAATTCAAGAGCTCGAATCATTC
TTTAAGGAATGTCCTCATCCAGATGATAAGCAACGAAAAGAGTTGAGCCGTGATCTCAAT
TTAGAGCCTCTTCAAGTTAAGTTTGGTTCCAAAACAACGCACACAGATGAAGGCACAA
AGTGAGAGGCATGAGAACCAGATTCTAAAGTCAGACAATGACAAGCTCAGAGCAGAGAAC
AATAGATACAAAGAAGCTCTAAGCAATGCTACATGCCCTAACTGTGGCGGTCCAGCTGCT
ATTGGAGAAATGCTTTTGTACGAACAACATCTCAGGATCGAAAATGCTCGGCTCCGCGAA
GAGATTGATAGGATCTCTGCTATTGCTGCGAAATACGTTGGGAAGCCGTTAGGATCGTCT
TTCGCTCCACTAGCGATCCACGCGCTTCTCGTTTCGTTGATCTTGAAGTTGGAACTTT
GGGAACCGACAGGCTTTGTAGGAGAAATGTATGGAACAGGGGACATTTTGAGGTCAAGTT
TCGATTCCTTCTGAGACTGATAAGCCTATAATCGTGGAGCTAGCGGTTGCAGCTATGGAG
GAACTCGTGAGAATGGCTCAAACCTGGAGATCCTTTATGGCTTCAACCGATAATTCAAGTC
GAGATTCTCAACGAAGAAGAGTATTTCAAGACGTTTCCGAGAGGAATTGGACCAAAGCCA
TTAGGATTAAAGATCAGAGGCGTCAAGACAATCTGCAGTTGTATAATGAATCACATCAAT
CTCGTTGAGATTCTCATGGATGTGAATCAATGGTCTTGTGTTTTCTCTGGGATTGTGTCA
AGAGCCTTGACACTTGAAGTTCTTTCAACTGGAGTTGCTGGGAACACAACGGTGCTTTA
CAAGTGATGACAGCTGAGTTTCAAGTTCATCACCCCTAGTCCCAACGCGTGAGAATAC
TTTGTGAGATACTGCAAACAACACAGTGACGGCTCTTGGGCTGTGGTTGATGTCTCTTTG
GACAGCCTTAGACCAAGTACTCCAATCTTAAGAACTAGAAGAAGGCCTTCAGGTTGTCTG
ATTCAAGAATTGCCTAATGGTTATTCTAAGGTTACATGGATAGAGCATATGGAGGTAGAT
GATAGATCAGTTTCAACATGTATAAACCGTTGGTTCAAGTCCGTTTAGCTTTTCGGTGCG
AAACGTTGGGTGGCTACACTCGAACGCAATGCGAGCGGCTTGCTAGCTCCATGGCCAGC
AACATTCTGGTGATCTTTCCGTGATAACGAGTCTGAAGGAAGGAAGAGTATGTTGAAG
CTAGCTGAGAGAATGGTTATGAGTTTCTGCAGTGGTGTGGCGCGTCACTGCACACGCT
TGGACAACAATGTGACAACAGGATCCGATGATGTTCCGGTCAAGAGTATG
GATGATCCAGGAAGACCTCCGGGTATTGTTCTTAGTGACGCTACTTCATTCTGGATCCCA
GTTGCTCCCAAACGTGTTTTTGTATTTCTCCGTGACGAAAATTCAAGAAAAGAGTGGGAT
ATTCTGTCAAATGGAGGTATGGTTTCAAGAAATGGCTCATATAGCCAATGGTCATGAACCT
GGAACTGTGTCTCCTTGTCTCCGAGTCAATAGTGGAACTCGAGCCAGAGCAACATGTTG
ATTCTACAAGAGAGCTGTACAGATGCATCAGGATCGTATGTGATTACGCGCCAGTGGAT
ATAGTGGCGATGAATGTGGTTCTAAGCGGTGGAGATCCTGATTACGTGGCGTTGTTGCCG
TCTGGTTTTGCTATTTTACCGGATGGTTCCGTTGGAGGAGGAGATGGGAATCAGCATCAG
GAAATGGTTTTCTACTACTTCTTCTGGGAGTTGTGGTGGTTTCGTTTTAACCGTTGCGTTT
CAGATTCTTGTGACTCTGTTCTTACAGCTAAACTCTCACTTGGCTCGGTGGCTACGGTT
AATAGTCTGATCAAATGTACGGTGGAGAGGATTAAAGCTGCTGTTTCTTGTGATGTTGGA
GGAGGAGCGTAG
>G1588 Amino Acid Sequence (domain in AA coordinates: 66-124)
MYHPNMFESHMFDMPKSTSDNDLGITGSREDDFETKSGTEVTTENPSGEELQDPSQRP
NKKKRYHRHTQRQIQELESFKECPHPDDKQKELSRDLNLEPLQVKFWFQNKRTQMKAO
SERHENQILKSDNDKLRAENNRYKEALSNAATCPNCGGPAAIGEMSFDEQHLRIENARLRE
EIDRISAIAAKYVGKPLGSSFAPLAIHAPSRSLDLEVGNGFNQTGFVGYGTGDIILRSV
SIPSETDKPIIIVELAVAAMEELVRMAQTGDPLWLSTDNSVEILNEEYFRFFRGIGPKP
LGLRSEASRQSAVIMNHINLVEILMDVNQWSCVFGIVSRALTLEVLSTGVAGNYNGAL
QVMTAEFQVPSPLVPTRENYFVRYCKQHSWVVDVSLDLRPSPTPILRTRRRRPSGCL
IQELPNQYKVTWIEHMEVDDRSVHNMYKPLVQSGLAFAKRWVATLERQERLASSMAS
NIPGDLSVITSPEGRKSMKLAEKRMVMSFCSGVGASTAHAWTTMSTTGSDDVVMTRKSM
DDPGRPPGIVLSAATSFWIPVAPKRVDFDLRDENSRKEWDILSNGGMVQEMAHIANGHEP

GNCVSLLRVNSGNSSQSNMLILQESCTDASGSYVIYAPVDIVAMNVVLSGGDPDYVALLP
SGFAILPDGSGVGGDGNQHQEMVSTTSSGSCGGSLLTVAFQILVDSVPTAKLSLGSVATV
NSLIKCTVERIKAAVSCDVGGGA*

>G1589 (179..2221)

ACCAAACCTACATAGCAATCACACACATCTCCACAAACACAGCTTGAGATGATCATGAAA
CACGTGCATCCTCAGATCTCTATCAATCCAGCTTGGTGAAAGAAGGTCAAGAATTGAAAG
AGAATCAAAGAAAACGACGTCGTTTCATTTCGTGTGTAACTACTAATTATACATAGAT
GGCTGCTTACTTTACGGAAACCCACCGGAGATCTCTGCCGGATCCGACGGTGGTCTTCA
AACGTTGATCCTCATGAATCCAATACTTACGTTTACGTACACCCAACAAGACAACGACTC
GAACAACAACAACAACAGCAACAATAGCAACAACAACAACAACAACAACAACAACA
CAACAACAGTAGTTTCGTTTTCCTCGATTCCCACGCGCCGACGCAAAACGCGAGCCAGCA
GTTTCGTCGGAATACCACTCTCAGGTACGAAGCTGCTTCCATTACAGCCGCCGACAACAT
CTCCGTACTTCACGGTTATCCTCCGCGCGTGCAGTACAGTCTCTACGGTAGCCACCAAGT
GGATCCCACTCACCAGCAAGCCGCGTGTGAGACGCCACGCGCGCAGCAAGGCCTCTCTTT
AACCCTCTCGTCTCAACAGCAGCAGCAACAGCAACATCATCAACAACACCAGCCTATTCA
CGTCGGATTCCGGGTCGGACATGGAGAAGATATCCGGGTCCGGTCTGGCTCTACAGGATC
GGGGGTAACAACCGGTATAGCTAATCTTGTAGCTCCAAGTACTTGAAGGCAGCACAAGA
GCTTCTTGACGAAGTAGTCAACGCTGATTCCGATGACATGAACGCTAAATCCCACTATT
CTCATCGAAAAAGGGTAGTTGCGGAAATGATAAACCTGTGCGAGAATCATCGGCCGGCGC
TGGAGGAGAAGGTTCCGGTGGCGGAGCAGAAGCAGCCGGGAAACGTCCGGTGGAGCTAGG
CACGGCAGAGAGACAAGAAATACAGATGAAGAAAGCAAACTTAGTAACATGCTTCATGA
GGTGGAGCAGAGATATAGACAGTACCACCAGCAGATGCAGATGGTGATCTCTTCGTTCTGA
GCAAGCGGCAGGGATAGGATCAGCGAAGTCATACACGTCGCTAGCATTGAAAACCATATC
AAGACAGTTCGTTGCTTGAAAGAGGCGATCGCTGGTCAGATAAAAGCGGCCAACAAAGAG
TCTTGGGGAGGAAGATTTCAGTGCTGTTGTTGGGAGGTTTGAGGGTTCGAGGCTCAAGTT
CGTGGACCACCACTTGAGACAGCAAAAGAGCTCTTCAACAACGGGAATGATTCAACATCC
TTCCAATAATGCTTGAGACCTCAACGTGGTCTCCAGAACGAGCCGTCTCAGTTCTCCG
TGCTTGGCTCTTCGAACACTTTCTTCATCCATACCCTAAGGATTCCGACAAGCACATGCT
AGCTAAGCAAACAGGACTCACTCGTAGCCAGGTGTGCAACTGGTTTATAAACCGCGAGAGT
TCGGTTATGGAACCAATGGTGGAGGAGATGTACATGGAGGAATGAAGGAGCAGGCAAA
GAACATGGGATCCATGGAAAAGACTCCTTTGGATCAAAGCAACGAAGATTCTGCTTCAA
GTCAACAAGTAACCAAGAAAAGAGCCCAATGGCGGACACTAATTACCATATGAATCCCAA
TCACAACCGGTACCTAGAAGCGCTCACTGGAATGCAAGGATGCCCAAGAGACTAAGAAC
CAGCGACGAGACAATGATGCGCAATATAATGCGGATTTTACGCTCCAACGAGAAGCTCAC
GATGAAGATTCTAGAAGAACGGCAAGGGATAAGATCAGATGGTGGCTACCTTTTCATGGG
TAATTTTCGGGCAATACCAATGGATGAGATGTCAAGATTTGATGTAGTCTCAGACCAGGA
GCTCATGGCGCAAAGGTACTCAGGAACAACAATGGCGTGTCCCTCACGTTAGGTTTACC
TCATTTGTGATAGCTTGTCTCCACGGACCATCAGGGTTTCATGCAGACCCACCATGGGAT
TCCTATAGGGAGAAGAGTGAAAAATAGGAGAAACAGAGGAATATGGACCCGCCACCATCAA
TGGTGGTAGCTCGACCACAACCGCACATTATCAGCGGCAGCTGCCGCGGCTTACAATGG
GATGAACATACAGAACCAGAAGAGATATGTGGCTCAGTTATTGCCCGACTTCGTTGCATA
AACCCATCTCTTAGAAGGAGAAACCGAAACAGGTTATTATATACGTTTCTAGTTTTTAA
TTAGTATATAGTTTTCTCATACCATTGAACCAAAACAAAGAAATTTAATTTTAGTCT
TTGGTTATATATGGCCGACGGGTACGTACAGGGCCCTGACGTAGC

>G1589 Amino Acid Sequence (conserved domain in AA coordinates:384-448)

MAAYFHGNPPEISAGSDGGLQTLILMNPTTYVQYTQDNDNSNNNNNSNNNNNTNTNTN
NNNSSFVFLDASHAPQPNASQQFVGIPLSGHEAASITAADNISVLHGYPYPRVQYSLYGSHQ
VDPETHQQAACETPRAQQGLSLTLSSQQQQQQHQQHQPPIHVFGSGHGEDIIVGSGSTG
SGVTNGIANLVSSSKYLKAAQELLDEVVNADSDDMNAKSQLFSSKKGSCGNDKPVGESSAG
AGGEGSGGGAEEAGKRPVELGTAERQEIQMKKAKLSNMLHEVEQRYRQYHQMQMVISSF
EQAAGIGSAKSYTSLALKTISRQFRCLKEAIAQIKANKSLGEEDSVSGVGRFEGSRLK
FVDHHLRQQRALQQLGMIQHPSNNAWRPQRGLPERAVSVLRWLFEHFLHPYPKDSKHM
LAKQTGLTRSQVSNWFINARVRLWKPMVEEMMEEMKEQAKNMGSMKTPLDQSNEDSAS
KSTSNQEKSPMADTNYHMNPNHNGDLEGVTGMQGPCPKRLRTSDETMMPINADFSSNEKL
TMKILEERQGIKSDGGYFPFGNFQYQMDMSRFDVVSQELMAQRYSGNNNGVSLTLGL
PHCDLSLSTDHQGFMTTHGIPIGRRVKIGETEEYGPATINGGSSTTTAHSSAAAAAAYN
GMNIQONQKRYVAQLLPDFVA*

>G160 (38..784)

TCAAATTTGTCATTGTGTTATTCAAATTTTGGAGAAAATGGTGAGAAAGTACCAAAGGTCG
TCAGAAAATAGAGATGAAAAAATGGAAAACGAAAGCAACCTTCAGGTTACTTTCTCAAA
AAGAAGATTCCGGTCTTTTCAAAAAAGCTAGTGAACCTTGCACATTAAGTGGTGCAGAGAT
TCTGTTGATTGTGTTCTCTCCTGGTGGGAAAGTGTTCCTTTGGCCATCCAAGTGTTC
AGAAGTCAATTCATCGCTTTTCGAATCCTAACCATTAATCTGCCATTGTCCATCATCAGAA
CAACAATCTCCAACCTGTTTGAACCCGTCGCGATAGAAAATCCAATATCTCAACAATAT
ACTCACTGAGGTGCTGGCAAACAGGAAAAGGAGAAACAGAAGAGAATGGTTTGGACCT
ATTGAAAGAATCCAGAGAACAGTAGGAACTGGTATGAAAAGATGTGAAAGATCTCGA
CATGAATGAAACCAACAGCTGATATCTGCTCTTCAAGATGTGAAAAGAAAAGTGGTAAG
AGAAATGTCTCAATATCTCAAGTAAATGTTTCGCAGAATTACTTTGGTCAAAGTTCTGG
CGTGATTGGTGGTGGTAAATGTTGGCATTGATCTTTTGGATCAAAGAAGAAATGCATTCAA
CTATAATCCAAACATGGTGTTCCTCAATCATACACCACCAATGTTTGGATACAACATGA
TGGAGTTCCTGTTCCGATATCCAACATGAATACATGTCAAGTTACAACCTCAACCAGAG
CTAGAGTCTGAAGCTAGAAGAACATCCTAATCAATATTTGCGTTATTTGGCTATGGTTA
CTGTTAGGATTGTTCTGTTATGTTGAGACTTAAGTTTGTTCCTTTTAATTTGTTTCA
GTTGGTTGGTTTTTTCATTTTATTTCGTCGTTTGTTCCTTTGTTTGGATATTTTGT
TCCCAGAATAAATTTATTTATCCTTTAAAAA

>G160 Amino Acid Sequence (domain in AA coordinates: 7-62)

MVRSTKGRQKIEMKKMENESNLQVTFKRRFLFKKASELCTLSGAEILLIVFSPGGKVF
SFGHPSVQELIHRFSNPNHNSAIVHHQNNNLQLVETRPDRNIQYLNILTEVLANQEKEK
QKRMVLDLLKESREQVGNWYEKDVKDLDMNETNQLISALQDVKKLVREMSQYSQVNVSQ
NYFGQSSGVIGGGNVGIDLDFQRRNAFNYPNMFVFNHTPPMFGYNNDGVLPVISNMNYM
SSYNFNQS*

>G1636 (19..666)

GAGTAATCATCAACGATTATGGCGTCAAGTCAGTGGACGAGGTCGGAGGATAAGATGTTT
GAGCAAGCTTTGGTCTTTTCTCTGAAGGATCTCCTAATCGGTGGGAGAGAATCGCTGAT
CAGCTTCATAAATCTGCTGGTGAAGTTAGGGAGCATTACGAGGTCCTGGTTCATGATGTT
TTCGAGATTGATTCTGGTTCGAGTTGATGTCCCTGATTACATGGATGACTCGGCGGCTGCG
GCGGCGGGTTGGGATTCCGCTGGTCAGATCTCTTTGGGTCTAAACATGGCGAGAGTGAA
CGCAAAGAGGAACCTCTTGGACAGAGAACGAACACAAATTGTTTCTGATCGGATTAAAG
AGATATGGTAAGGGAGATTGGAGGAGTATCTCGAGAAACGTTGTGGTGACGAGGACACCG
ACGCAAGTCGCGAGTACGCTCAGAAGTATTTTCTGAGACAGAACTCGGTGAAGAAGGAG
AGGAAAAGGTTCGAGCATCCATGATATAACTACGTTGATGCTACTTTGGCTATGCCTGGG
TCTAACATGGACTGGACTGGCCAAACGCGGAGTCTGTTAGGCGCCGAGCAGCAACAG
ATTATGTCTGAGTTCGGTCAGCAATTGAATCCTGGTCATTTGAGGATTTGGGTTTCGG
ATGTGATG

>G1636 Amino Acid Sequence (domain in AA coordinates: 100-165)

MASSQWTRSEDKMFEQALVLFPEGSPNRWERIADQLHKSAGEVREHYEVLVHDVFEIDSG
RVDVPDYMDDSAAGWDSAGQISFGSKHGESERKRGTPTWENEHKLFLIGLKRYGKD
WRSISRNVVVTRPTQVASHAQKYFLRQNSVKKERKRSSIHDTTVDATLAMPGSNMDWT
GQHGSFPVQAPQQQIMSEFGQQLNPGHFEDFGFRM*

>G1642 (1..1077)

ATGGGTCATCACTCATGCTGCAACAAGCAAAAGGTGAAGAGAGGGCTTTGGTCACCTGAA
GAAGACGAAAAGCTCATCAACTACATCAATTCATATGGCCATGGATGTTGGAGCTCTGTT
CCTAAACATGCAGGTTTGCAGAGATGTGGAAAGAGTTGTAGATTAAAGATGGATAAATTAT
CTAAGACCTGATCTTAAACGTGGAAGCTTCTCTCCTCAAGAAGCTGCTCTTATCATTGAG
CTTCACAGCATCTTGGTAACAGATGGGCTCAAATTGCTAAACATCTACCTGGAAGAACA
GATAACGAGGTCAAGAATTTCTGGAACCTCGAGCATTAAAAAGAAGCTCATGTCTACCAT
CATCACGTCATCATCATCATCTCTCTTCATGGCGAGTTTGCTCACAAACCTTCTCT
TATCACAATGGATTCAACCTTACTACAGTGCAGATGAAAGTTCAAGATTCATGTCCAAT
ATCATCACAAACACTAACCTAATTTTCATCACTCCAAGCCATCTCTCTCTCTCTCTCTCT
CATGTTATGACCCATTGATGTTCCCAACCTCTAGAGAAGGAGATTTCAAGTTTCTAACCC
ACAAACAACCCAAACCAATCTCATCACCATGATAATAACCATTACAACAACCTCGACATT
TTGTACCCACACCAACTATAACAATCATCATCAACCTTCACTTTCTTCTTGTCTCAT
GATAATAATCTCCAATGGCCAGCGTTACCAGATTTCCAGCGAGTACCATTTCTGGTTTC
CAAGAAACCTTCAAGATTATGATGATGCTAATAAACTCAACGTGTTTGTGACACCATT

AACGATAATGCCAAAAAGTTATTATGTGGAGAAGTTCTCGAAGGCAAAGTACTATCTTCC
 TCCTCACCAATTTCAACAAGATCACGGCCTTTTCTTCCCACCACGTACAACCTTTCAAATG
 ACTTCTACGAGTGATCATCAACATCATCATCGAGTGGACTCATACATCAATCACATGATC
 ATACCATCATCATCCTCATCGTCGCCAATCTCTGTGGACAGTACGTCATAACTTAA

>G1642 Amino Acid Sequence (domain in AA coordinates: TBD)

MGHHSCCNKQKVKRGLWSPEEDEKLINYINSYGHGCWSSVPKHAGLQRCGKSCRLRWINY
 LRPDLKRGFSFSPQEAALIELHSILGNRWAQIAKHLPGRTDNEVKNFVNSSIKKLMSSH
 HHGHHHHHLSSMASLLTNLPYHNGFNPTTVDESSRFMSNIITNTPNFITPSHLSLPSP
 HVMTPLMFPTSREGDFKFLTNNPNQSHHHDNNHYNLDILSPTPTINNHHQPSLSLSSCPH
 DNNLQWPALPDFPASTISGFQETLQDYDDANKLNVFVTPFNDNAKLLCGEVLEGKVLSS
 SSPISQDHGLFLPTTYNFQMTSTSDHQHHRVDSYINHMIIPSSSSSSPISCGQYVIT*

>G1747 (1..777)

ATGAAAATGATGCAAGAGGAGGAAACCGAAAAGGTCCATGGACAGAACAGGAAGACATA
 CTTCTGGTAAATTTTGTTCACCTTATTTGGAGATCGACGATGGGATTTTATAGCAAAAGTA
 TCAGGTTTGAACAGAACAGGAAAGAGTTGCGAGCTAAGATGGGTTAATTACCTACATCCT
 GGTCTCAAACGTGGCAAGATGACGCCTCAAGAAGAGCGCCTCGTCTTGAGCTTCACGCT
 AAGTGGGGAAACAGGTGGTCGAAAATAGCCCCGAAAATGCGCGGACGAACGGATAACGAG
 ATAAAGAACTACTGGAGGACTCATATGAGGAAGAAAGCTCAAGAAAAGAAGCGTCTGT
 TCCCCAACCTTCTCATTTTCCAACCTGCAGCTCGTCATCTGTGACCACTACCACCACCAAT
 ACTCAAGATACATCGTGCCACTCGCGTAAATCTTCAGGGGAAGTGAGCTTTTACGACACT
 GGAGGTTCCCGATCCACTAGAGAGATGAATCAAGAAAACGAAGACGTGTACTCGTTGGAT
 GATATATGGAGAGAGATTGATCACTCAGCAGTAAACATAATAAAACCGGTTAAAGACATC
 TACTCAGAACAAAGCCATTGCTTAAAGTTACCCAAATCTAGCTTACCATCATGGGAAAGC
 TCATTGGATTCTATATGGAACATGGATGCAGATAAAAGTAAGATATCGTCTTACTTTGCA
 AATGATCAGTTTCTTTCTGTTTCCAACACAGTAGATCACCATGGTCGTCAGGTAA

>G1747 Amino Acid Sequence (domain in AA coordinates: 11-114)

MKMMQEEGNRKGPPWTEQEDILLVNFVHLFGDRRWDFIAKVSGLNRTGKSCRLRWVNYLHP
 GLKRKGMTPOEERLVLELHAKWGNRWSKIARKLPGRDNEIKNYWRTHMRKKAQEKRPV
 SPTSSFSNCSSSSVTTTTNTQDTSCHSRKSSGEVSFYDTGGSRTREMNQENEDVYSLD
 DIWREIDHSAVNIKPKVDIYSEQSHCLSYPNLASPSWESSLDSIWNMDADKSKISSYFA
 NDQFPFCFQHSRSPWSSG*

>G1749 (59..535)

CAACACTTCTCAGTGACCGTGAGCAACGAATTATTTTTCAGTTCAACGACTCCGCGGAAAT
 GGAAAATTGAGAAAATGTTCCCTCTTACGATCAAAACATCAATTTCACTCCTAATTGAC
 GAGAGATCAAGAACATGTGATCATGGTCTCTGCTTTGCAACAAGTAATATCCAACGTCGG
 AGGTGACACGAACTCGAATGCATGGGAAGCTGATCTTCCACCTTTGAACGCTGGCCCTTG
 TCCTCTTTGTAGTGTACCCGGCTGCTACGGTTGCGTCTTCCCACGACACGAGGCGATAAT
 TAAGAAGGAGAAGAAGCACAAAGGAGTGAGGAAAAAACCATCAGGTAAATGGGCGCGGA
 GATATGGGATCCGAGTTTGAAGTAAGGAGATGGCTTGAACGTTTCCAACAGCGGAGAT
 GCGCGCTAAGGCTTACAACGATGCGGCGCTGAGTTTGTGGAAGAAGATCAGCAAGACG
 TGGCACAAAGAAGCGGAGAGGAAGCATCTACCAAGAAGACGACTGAGAAAAATTAACGGAG
 AAGGAGCACGTATAGAAAGGCAGGAAGAGGCATCTTACTTGCTTCACAAGTAAATCAGAA
 TTTTTTTGAAAAGTAAAAACGTTATTTTGTGTTGGTAATAAAATAAAGTAAAAACAAATAT
 TGCTAACGCAAGACTTATCAAGTTCAAGTCGTGACTGTGAGTGTGTTTTATGTATCTTAC
 TTCATTTTTTGTCTTTCAATTGTGTGTGTGTGT

>G1749 Amino Acid Sequence (conserved domain in AA coordinates: 84-155)

MENSENVPSYDQNIPTFNLTRDQEHVIMVSALQQVISNVGGDTNSNAWEADLPPLNAGP
 CPLCSVTGCGYCVFPRHEALIKKEKKHGVRRKPSGKWAABEIWDPSLKVRRLGTFPTAE
 MAAKAYNDAAEFVGRRSARRGTKNGEASTKKTEKN*

>G1751 (117..923)

AAACACAAACAAAACCTCATATTTTCAATCTCCAGGTGCTTTACACCAACAGAGTCGCAAG
 AAAACAAAACCAAACCTCGGATTTAGTTTGACAGAAGAAGGAATCGAGAGTCGGGTATGC
 ATTATCCTAACAAACAGAACCGAATTCGTGCGAGCTCCAGCCCCAACCCGGTATCAAAAGG
 AGCAGTTGTACCGGAGCAAGAGCTTTCAGTTATTGTCTCTGCTTTGCAACACGTGATCT
 CAGGGGAAAACGAAACGGCGCGGTGTGAGGTTTTTCAGTGACAGCACAGTGATAAGCG
 CGGGAATGCCTCGGTTGGATTCAGACACTTGTCAAGTCTGTAGGATCGAAGGATGTCTCG
 GCTGTAACACTTTTTTCGCGCCAAATCAGAGAATTGAAAAGAATCATCAACAAGAAGAAG

AGATTACTAGTAGTAGTAACAGAAGAAGAGAGAGCTCTCCCGTGGCGAAGAAAGCGGAAG
GTGGCGGGAAAAATCAGGAAGAGGAAGAACAAGAAGAATGGTTACAGAGGAGTTAGGCAAA
GACCTTGGGGAAAAATTTGCAGCTGAGATCAGAGATCCTAAAAGAGCCACACGTGTTTGGC
TTGGTACTTTTCGAAACCGCCGAAGATGCGGCTCGAGCTTATGATCGAGCCGCGATTGGAT
TCCGTGGGCCAAGGGCTAAACTCAACTTCCCCTTTGTGGATTACACGTCTTCAGTTTCAT
CTCCTGTTGCTGCTGATGATATAGGAGCAAAGGCAAGTGCAAGCGCCAGTGTGAGCGCCA
CAGATTCAAGTTGAAGCAGAGCAATGGAACGGAGGAGGAGGGGATTGCAATATGGAGGAGT
GGATGAATATGATGATGATGATGGATTGTTGGGAATGGAGATTCTTCAGATTCAAGAAATA
CAATTGCTGATATGTTCCAGTGATAAATGAGCTCTTTCTTGTGGCGTTTTTTGGAGTTA
AGTGCAAGAAGAGATTGACACTGTGGCTTGTTTAAAGTGAACAAGAAAGCAATGT
AATTAGTAGTCTCATCTTTTGTGTTGTGGTCAATTCTATGTTTATCTCATATAAAATCTG
AGTTAAACCTATCTGAGGAGAGAGTAAATAAAGAGGTTAAGAA

>G1751 Amino Acid Sequence (domain in AA coordinates: TBD)
MHYPNNRTEFVGAPAPTRYQKEQLSPEQELSVIVSALQHVISGENETAPCQGFSSDSTVI
SAGMPRLSDSTCQVCRIEGLGCNYFFAPNQRIEKNHQEEIITSSNNRRRESSPVAKKA
EGGGKIRKRKNKNGYRGVRQRPWGKFAAEIRDPKRATRVWLGTFFETAEDAAARAYDRAAI
GFRGPRAKLNFPFVDYTSVSSPVAADDIGAKASASVSATDSVEAEQWNGGGGDCNME
EWMNMMMMDFNGDSSDSGNTIADMFG*

>G1752 (25..756)

AAAAAAAAAAAAAAAAAACTTATGGAATATTCCCAATCTTCCATGTATTTCATCTCCA
AGTTCTTGGAGCTCATACAAGAATCACTCTTATGGAACGAGAGCTGTTTCTTGGATCAA
TCATCTGAACCTCAAGCCTTCTTTTGCCCTAATTATGATTACTCCGATGACTTTTCTCA
TTTGAGTCACCGGAGATGATGATTAAAGGAAGAAATTCAAACCGGCGACGTTTCTAACTCC
GAAGAAGAAGAAAAGGTTGGAATTGATGAAGAAAGATCATAAGAGAGTGAGGAAAAGG
CCGTGGGGGAAATTTGCAGCGGAGATAAGAGATTCAACGAGGAATGGAATTAGGGTTTGG
CTCGGGACATTTGACAAAGCCGAGGAAGCCGCTCTTGCTTATGATCAAGCGGCTTTCGCC
ACAAAAGGATCTCTTGCAACACTTAATTTCCCGGTGGAAGTGGTTAGAGAGTCGCTAAAG
AAAATGGAGAAATGTGAATCTTCATGATGGAGGATCTCCGGTTATGGCCTTGAAGAGAAAA
CATTCTCTTCGAAACCGGCCTAGAGGGAAAAAGCGATCCTCTTCTTCTTCTTCTTCT
TCTAATCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AAGCAGAGTGTGTGAAGCAAGAAAGTGGTACACTTGTGGTTTTTGAAGATTAGGTGCT
GAGTATTTAGAACAACCTTCTTATGAGCTCATGTTGATCTTGTAATTGATTTCAGCAAAAG
CCACTATTAACCTTTAATTTTGTGATAATTAATCTTGAAATTTGTTTTGTTCTTCTGCA
ATTTCTTTGGTTCTCTTATTTTTTGTGTTGTGATCCAAATGAAATTATTGGAAGAGATG
GTGATGTTAAAGTGATATATATATAAAAAAAAAA

>G1752 Amino Acid Sequence (domain in AA coordinates: TBD)
MEYSQSSMYSSPSSWSSSQESLLWNESCFLDQSSEPQAFPCPNYDSDFFSFESPEMMI
KEEIQNGDVSNSEEEKVGIDEERSYRGVRKRPWGKFAAEIRDSTRNGIRVWLGTDFDKE
EALAYDQAAFPATKSLATLNFPEVVRESLKKMENVNLHDGGSPPMALKRKHSRLNRPR
GKKRSSSSSSSSSSSSSSSSSTSSSTRSSSKQSVVKQESGTLVVFEDLGAEYLEQLLM
SSC*

>G1763 (33..977)

GTACATTTTTTTTGTATTTTCAAGAACTCCGATGGCGGATCTCTTCCGGTGGTGGCCACG
GCGGCGAGCTTATGGAAGCACTTCAACCTTTTTTACAAAAGTGCTTCCACGTCTGCTTCAA
ATCCTGCGTTTGCCTCTCAAACGATGCGTTTGCCTGCGCCCAAACGACCCATTTCTT
CTTCTTCTTACTATAATCCTCATGCATCTTCTTCCCTTCAATTCCACAACCACTTACC
CGGATATTTATTTCTGGATCCATGACCTATCCATCTTCACTCGGGTCCGATCTTCAACAAC
CCGAAAACCTACCAATCTCAGTTCCATTACCAAAACACTATCACTTCACTCACCAGACA
ACAACACTTGCGATGCTCAACTTCATTGAGCCGAGCCAACCGGATTTTATGACCCAACCGG
GTCCGAGTTCCGGTTEGGTTTCAAAACCGGCTAAGCTCTATAGAGGAGTGAGGCAAAGAC
ATTGGGGAAAATGGGTGCGCGAGATCCGTTTACCCAGGAACCGAACCCGACTTTGGCTCG
GAACATTCGACACGGCTGAAGAAGCCGCGTTGGCTTATGATCGCGCCGCTTTAAGCTTC
GTGGTGAATCGGCTCGGCTTAACTTCCCAGCTCTCCGATACCAACCGGCTCGTCTCCGT
CTGACGTTGGCGAATACGGACCTATTCAAGCTGCCGTTGACGCCAAGCTAGAAGCCATAT
TAGCTGAGCCGAAGAATCAGCCGGGCAAAACGGAGAGGACGTCGAGGAAACGAGCTAAAG
CCGCGGCTTCTTTCAGCTGAGCAGCCGTCAGCGCCACAACAACATTCGGGTTCGGGTGAA
GTGATGGGTCCGGTTACCGACTTCGGATGTTATGGTGCAGGAGATGTGCCAAGAGCCAG

AGATGCCATGGAATGAAAATTTTCATGCTCGGCAAGTGTCTTCTTATGAGATAGATTGGG
 CTTCAATTTTATCGTGAAAAATTAGGATTCATTTTATTCATTTTAACTTGTTTG
 TATTTTCTTTAACTTTAGGGTTATTAGCTGTGCGTAAATTTGTAATTTAGCATTTTG
 TATGAATGTAATGCAAGTGTGTAAATTATGGACAGCTCAAGCTTTTTTGTAAAA

>G1763 Amino Acid Sequence (conserved domain in AA coordinates:140-209)

MADLFGGGHGGELMEALQPFYKSASTSASNPFASSNDAFASAPNDPFSSSSYNPHASF
 FPSHSTTTYPDYSGSMTYPSFSGDLQQPENYQSQFHYQNTITYTHQDNNTCMLNFI
 EQPDMFTQPGPSSGSVSKPAKLYRGVQRHGWKVAEIRLPRNRTRLWLGTFTDTEEAAL
 AYDRAAFKLGRDSARLNFPALRYQTGSSPSDVGEYGPIQAAVDKLEAILAEPKNQPGKT
 ERTSRKRAKAAASSAEQPSAPQQHSGSGESDGSPTSDDVMVQEMCQEPMPWNENFMLG
 KCPSYEIDWASILS*

>G1766 (32..1216)

AGGCTATTCTCGAAAAACAAGAATAAAGAATGAATTCGTTTTCACAAGTACCTCCTGG
 CTTTCAGATTTTCATCCTACTGATGAAGAACTTGTAGACTACTACTTGAGGAAAAAGTTGC
 ATCAAAGAGAATAGAAATCGATATCATCAAGGATGTTGATCTTTACAAGATTGAGCCATG
 TGATCTTCAAGAGTTATGCAAGATAGGAAACGAAGAGCAGAGCGAATGGTACTTCTTTAG
 TCATAAAGACAAGAGTATCCCACGGGAACTCGAACCAATAGAGCCACGAAAGCAGGATT
 TTGGAAAGCCACTGGAAGAGACAAGGCATATATATAAGACATAGTCTTATCGGTATGAG
 GAAACACTTGTGTTTTACAAAGGAAGAGCCCCAAATGGTCAGAAATCCGATTGGATCAT
 GCACGAATATCGCTTAGAAACAAGTGAATGGAACCCCTCAGGAAGAAGGATGGGTAGT
 ATGTAGGGTATTCAAGAAGAAATTTGGCAGCGACAGTGAGGAAATGGGAGATTACCATTC
 ATCACCATCGCAGCATTTGGTACGATGATCAGCTCTCTTTTATGGCCTCCGAGATCATTC
 TAGTTCTCCACGACAGTTTCTTCCAATCATCATTATAACCGCCACCATCACCAGCAGAC
 ATTGCCTTGTGGCCTCAATGCATTCAACAACAACATCCTAAGTGAATGCAAGCAAGA
 GCTCGAGTTACATTACAATCAAATGGTACAACATCAACAACAAACCATCATCTTCGTGA
 ATCTATGTTTCTCCAGCTTCCTCAGCTCGAAAGCCCTACCAGTAATTGCAATTCTGACAA
 CAACAATAACACAAGAAATATTAGTAAC'TTGCAAAATCATCAAATATATCTCATGAGGA
 ACAATTGCAACAAGGAATCAAAGTTTCACTCTCTGTATTACGATCAAGGAGTAGAGCA
 AATGACTACTGACTGGAGAGTTCTCGATAAATTTGTTGCTTCACAGCTTAGCAATGATGA
 AGAGGCTGCAGCCGTGGTTTCTTCTTCTTCTCATCAAAACAACGTCAAGATTGACACGAG
 AAACACGGGTTATCATGTGATAGATGAGGGAATAAATTTGCCGAGAAATGATTCTGAAAG
 GGTGTTGAAATGGGAGAAGAGTATTCAAATGCTCATGCTGCTTCTACTTCTTCAAGTTG
 TCAGATTGATCTCTAGAAATAGTGATAGAGAGATGAAAAAGATGCAAGGTGAATATATAT
 GAAATACATGCACACTAGTGTTATTTATACCTTAAAGATGGAAGGGGAAAAACAAGGAGT
 TATTTCTCGGATTTATGGAGGTTTTGTACATAATAAAACCTACAACCATATGGTATTTT
 CTTTTGAAAAAAAAAAAAAAAAAAAAA

>G1766 Amino Acid Sequence (domain in AA coordinates: 10-153)

MNSFSQVPPGFRFHPDTDEELVDYLLRKKVASKRIEIDIIKDVDLYKIEPCDLQELCKIGN
 EEQSEWYFFSHDKKYPGTGTRTNRAKAGFWKATGRDKAIYIRHSLIGMRKTLVFKGRA
 PNGQKSDWIMHEYRLTSENGTPQEEGWVVCVFKKLAATVRKMGDYHSSPSQHWYDDQ
 LSFMASEIISSSPRQFLPNHHYNRHHHQQTLPCLNLFNNNNPNLQCKQELHYNQMVQ
 HQQNNHHLRESMFLQLPQLESPTSNCSNDDNNNNTRNINSLQKSSNISHEEQLOQGNQSFS
 SLYYDQGVQMTTDDWRVLDKFWASQLSNDEEAAAVVSSSSHQNNVKIDTRNTGYHVIDEG
 INLPENDSERVEMGEEYSNAHAASTSSSCQIDL*

>G1767 (1..1596)

ATGGATACTCTCTTTAGACTAGTCAGTCTCCAACAACAACAACATCCGATAGTATCATT
 ACAAATCAATCTTCGTTAAGCAGAACTTCCACCACCACTACTGGCTCTCCACAACTGCT
 TATCACTACAACCTFFCCACAAAACGACGTCGTCGAAGAATGCTTCACTTTTTTCATGGAT
 GAAGAAGACCTTTCTCTTCTTCTTCTCACCACAACCATCACAACCACAACAATCCTAAT
 ACTTACTACTCTCTTTCACTACTCCCAACCAATACCATCCCGCCACATCATCAACCCCT
 TCCTCCACCGCGCGAGCCGAGCTTTAGCCTCGCCTTACTCCTCCTCCGGCCACCATAAT
 GACCTTCCGCGTTTCTCCATACCTCAAACCTCCTCGTCTCTCGACTTCTCAGCCAATGCC
 AAGTGGGCGAGACTCGGTCTTCTTGAAGCGGCACGTGCCTTCTCCGACAAAGACACTGCA
 CGTGCGCAACAAATCCTATGGACGCTCAACGAGCTCTTCTCCTCGTACGGAGACACCGAG
 CAAAACTGGCTTCTTACTTCTCTCAAGCTCTTCAACCGCATGACCGGTTCAAGCGAA
 CGATGCTACCGAACCATGGTAACAGCTGCAGCCACAGAGAAGACTTGCTCTTTCGAGTCA
 ACGCGAAAACTGTACTAAAGTTCCAAGAAGTTAGCCCTGGGCCACGTTTGGACACGTG

GCGGCAAACGGAGCAATCTTGAAGCAGTAGACGGAGAGGCAAAGATCCACATCGTTGAC
ATAAGCTCCACGTTTTGCACTCAATGGCCGACTCTTCTAGAAGCTTTAGCCACAAGATCA
GACGACACGCCTCACCTAAGGCTAACCACAGTTGTCGTGGCCAACAAGTTTGTCAACGAT
CAAACGCGCTCGCATCGGATGATGAAAGAGATCGGAAACCGAATGGAGAAATTCGCTAGG
CTTATGGGAGTTCTTTCAAATTTAACATTATTCATCACGTTGGAGATTATCTGAGTTT
GATCTCAACGAACCTCGACGTTAAACCAGACGAAGTCTTGGCCATTAAGTGCCTAGGCGCG
ATGCATGGGATCGCTTACGTTGGAAGCCCTAGAGACGCTGTGATATCGAGTTTCCGACGG
TTAAGACCGAGGATTGTGACGGTCGTAGAAGAAGAAGCTGATCTTGTGCGAGAAGAAGAA
GGTGGCTTTGATGATGAGTTCTTGAGAGGGTTTGGAGAATGTTTACGATGGTTTAGGGTT
TGCTTCGAGTCATGGGAAGAGAGTTTTCAGGACGAGCAACGAGAGGTTGATGCTAGAG
CGTGCACGGGACGTGCGATCGTTGATCTTGTGGCTTGTGAGCCGTCGGATTCCACGGAG
AGGCGAGAGACAGCGAGGAAGTGGTCGAGGAGGATGAGGAATAGTGGGTTTGGAGCGGTG
GGGTATAGTGATGAGGTGGCGGATGATGTCAGAGCTTTGTTGAGGAGATATAAAGAAGGT
GTTTGGTCGATGGTACAGTGTCTGATGCCGCCGAATATTCCTTTGTTGGAGAGATCAG
CCGGTGGTTTGGGCTAGTGGCTGGCGGCCAACGTAA

>G1767 Amino Acid Sequence (domain in AA coordinates: 255-272)

MDTLFRLVSLQQQQSDSIITNQSSLRTSTTTGSPQTAYHYNFPQNDVVEECFNFFMD
EEDLSSSSSHNHNHNHNPNNTYSPFTTPTQYHPATSSTPSSTAAAAALASPYSSSGHHN
DPSAFSIPQTPPSFDFSANAKWADSVLLEAARAFSDKDTARAQQILWTLNELSSPYGDTE
QKLASYFLQALFNRMGTSGGERCYRTMVTAAATEKTCSEFSTRKTVLKFQEVSPWATFGHV
AANGAILEAVDGEAKIHIVDISSTFCQWPTLLEALATRSDDTPhRLRTTVVVKFVND
QTASHRMKEIGNRMEKFARLMGVPFKFNIIHHVGDLSFDLNELDVKPDEVLAINCVA
MHGIASRGSPRDAVISSFRRLRPRIVTVVEEADLVGEEGGFDDEFRLRGFGECLRWFVRV
CFESWEESFPRTSNERLMLERAAGRAIVDLVACEPSDSTERRETARKWSRRMRNSGFGAV
GYSDEVADDVRALLRRYKEGVMSMVQCPDAAGIFLCWRDQPVVWASAWRPT*

>G1778 (1..627)

ATGATGGGATACCAAACTCTAATTTCTCCATGTTTTTTTCTCGGAAAATGACGAC
CAAAACCACCACAACACTACGATCCTTATAATAATTTCTCTTCATCAACTTCTGTTGATTGC
ACTCTCTCACTTGGAAACCCCTACTCGTCTCGACGACCACCATAGATTTTCTTCTGCT
AATTCTAACAACATCTCCGGCGACTTTTATATTACGAGGAAACGCTAAGACTTCTTCG
TACAAGAAGGGTGGTGTGCTCATAGCCTACCTCGCCGTTGTGCTAGCTGCGACACCACT
TCAACTCCTCTATGGAGAAACGGACCAAAAGGACCTAAGTCGTTATGTAACGCGTGTGGA
ATCCGATTCAAGAAAGAGGAGAGGCGTGGCAGCGCCAGAACTTAACGATCTCCGGTGGG
GGTTCATCAGCGGCAGAGTCCAGTAGAGAATTCGTACAACGAGGTGGAACTATTAC
AGTCATCATCATCACTATGCCTCGTCCGTCGTCGCGCTCATCAGAACACACAA
AGAGTTCCATATTTCTACCGGTTCCGGAGATGGAATATCCCTACGTGGATAACGTCACG
GCTTCTTCTTTTATGCTCTTGAATTGA

>G1778 Amino Acid Sequence (domain in AA coordinates: 94-119)

MMGYQTNSNFSMFFSSSENDQNHNYDPYNNFSSSTSVDCTLSLGPSTRLDHHRFSSA
NSNNISGDFYIHGNAKTSSYKGGVAHSLPRRCASCDTTSTPLWRNGPKGPKSLCNACG
IRFKKEERRATARNLTISGGGSSAAEVPVENSYNGGNYSHHHHHYASSSPSWAHQNTQ
RVPYFSPVPEMEYPYVDNVTASSFMSWN*

>G1789 (108..413)

CAAGGACTCTGCGACATCTGTGCAACATATCATTTCTCAGAATCTCTTTCTTTCTAGG
TTTATTACTACACAAACCAACATCATCAACTTTAGTTACTAAACAATGGCATCAGGCT
CAATGTCTTCTTATGGCTCTGGCTCATGGACTGTTAAGCAGAACAAAGCCTTTGAGCGTG
CTCTAGCAGTCTATGACCAAGACACTCCGGACCGTTGGCACAATGTTGCTAGAGCTGTTG
GTGGTAAACACCAGAGAAGCTAAGAGACAGTATGACCTTCTAGTTCTGTGACATCGAAA
GCATCGAGAATGGTCACGTGCCATTCCCTGACTACAAGACTACTACAGGAAACAGCAACA
GAGGCAGGCTGCGTGATGAGGAAAAGAGGATGAGAAGCATGAAGCTGCAGTGAGACAAGA
AGCAACAAAACCTAACTACGTATGATCGTCAAAATAAAAGAGAATCACTTCAGAGAGATG
TGTTTTTTTCAATGTCTGACGAATCAATGTTTTTTTCTTGCAATTTCTCATGTTTTTCCC
TAAGAAATGGTTTTTTTTCGAGGCAACAAAAA

>G1789 Amino Acid Sequence (domain in AA coordinates: 1-50)

MASGSMSSYSGSWTVKQNKAFERALAVYDQDTPDRWHNVARAVGGKTPPEAKRQYDLLV
RDIESIENGHVFPDYKTTTGNNSNRGLRDEEKRMRSMKLQ*

>G1790 (63..1346)

GAAAAAGACTTCACTTTTTTTTTTTACTAATTAATTAGTTTTTTTTTCTCCTTTCCAAAA
 CAATGGAGAATTTTCGTCGACGAGAATGGTTTTGCTTCTCTAAACCAAAACATCTTCACAC
 GTGATCAAGAACACATGAAAGAAGAAGATTTTCCATTTCGAAGTCGTCGACCAATCAAAAC
 CTACAAGCTTTCTTCAAGATTTTCACCATCTTGATCATGATCATCAGTTTGATCATCATC
 ATCATCATGGCTCCTCATCTTCACATCCTTTGCTCAGCGTCCAAACTACGTCTTCTTGT
 TCAATAATGCTCCTTTTCGAGCATTGCTCTTACCAAGAAACATGGTCGATTTCTATGAAA
 CTAACCAAAATTTGATGAATCATCATCATTTTCCAAGCAGTGGAAAACATCACTTCACTC
 GTAATCATCATCATCATCAAGAGATCAATTTGGTCGATGAACATGATGATCCTATGGACT
 TGGAGCAAAACAACATGATGATGATGAGGATGATCCCTTTTGATTACCCCTCTACAGAGA
 CTTTCAAACCTATGAACTTCGTAATGCCAGATGAAATTTTCATGTGTTTCTGCAGATAATG
 ATTTGTTATAGAGCAACGAGTTTCAACAAGACCAACCATTTCTTACACGAAAGTTGTCTT
 CTTCTTCTTCATCATCATCATGGAAGAAACCAAAAAGTCAACCTTAGTCAAAGGACAAT
 GGACTGCTGAAGAAGACAGGGTACTGATTCAACTCGTGGAGAAGTATGGATTGCGTAAAT
 GGTCGCATATCGCTCAAGTGTACCGGGAAGAATCGGGAACAATGTAGAGAGAGGTGGC
 ATAACCATTTGAGACCTGACATTAAAGAAAGAAACATGGAGTGAAGAAGAGGACAGAGTGT
 TGATAGAATTTCACAAAGAGATTGGAAACAAATGGGCAGAGATTGCGAAAAGACTCCCGG
 GAAGAACAGAGAAGTTCGATCAAGAACCATTGGAACGCAACAAAAGAAGACAATTCTCTA
 AAAGAAAATGTAGATCTAAGTATCCAAGACCTTCTCTGTTGCAGGATTACATCAAGAGCT
 TGAATATGGGAGCTTTGATGGCTTCTTCTGTTCTGCAAGAGGTAGACGCAGAGAGAGTA
 ATAACAAGAAGAAGGATGTTGTTGTTGCGGTTGAGGAGAAGAAGAAGGAAGAGGAGGTGT
 ATGGACAAGACAGGATTGTGCCTGAATGTGTGTTTACTGATGATTTTGGATTCAATGAGA
 AGCTGCTTGAGGAAGGATGTAGCATTGACTCTTTGCTTGATGACATTCCTCAGCCTGACA
 TTGATGCTTTTGTGTTTCTGATTTGATTTGATTTTATTCTGCTTGTTCAGTTTTGT
 TGTTTTTTGTGTTGCTTTTTTATACGAGACAGATTCCACCAAACTTCAATAATTTGAAAAG
 ATATAAAATATTTTGTCTTTTTAAAAAAAAAAAAAAAAAAAAAAAAA

>G1790 Amino Acid Sequence (conserved domain in AA coordinates:217-316)

MENFVDENGFASLNQNIIFTRDQEHMKEEDFPFEVVDQSKPTSFLQDFHHLDDHDHDFDHHH
 HHGSSSSHPLLSVQTTSSCINNAPFEHCSYQENMVDYETKPNLMNHHHFQAVENSFYFTR
 NHHHHQEINLVDEHDDPMDLEQNMMMMRMIPFDYPPTETFKPMNFVMPDEISCVSADND
 CYRATSFNKTKPFLLTRKLSSSSSSSSWKETKKSTLVKGQWTAEDRVLIQLVEKYGLRKW
 SHIAQVLPGRIGKQCRERWHNHLRPDIKKETWSEEDRVLIEFHKEIGNKWAEIAKRLPG
 RTEANSIKNHNWATKRRQFSKRKCRSKYPRPSLLQDYIKSLNMGALMASSVPARGRRRESN
 NKKKDVVVAVEEKKKEEVYQDRIVPECVFTDDPGFNEKLLLEEGCSIDSLLDDIPQPD
 I DAFVHGL*

>G1791 (36..455)

ATGTACATGCAAAAAACAAAACCTTAAAGCTTTTCATGGAAACGTATAGAGTCTTATAACA
 CGAATGAGATGAAATACAGAGGCGTACGAAAGCGTCCATGGGGAAAATATGCGGCGGAGA
 TTCGCGACTCAGCTAGACACGGTGCTCGTGTTTGGCTTTGGGACGTTTAAACACAGCGGAAG
 ACGCGGCTCGGGCTTATGATAGAGCAGCTTTCGGCATGAGAGGCCAAAGGGCCATTCTCA
 ATTTTCTCAGCGATTCAAATGATGAAGACGGTCCAAATGGCAGCCACGAGAATGCAG
 TGGCTTCTCTCGTCTCGGGATATAGAGGAGGAGGTGGTGGTGATGATGGGAGGGAAGTTA
 TTGAGTTTCGAGTATTTGGATGATAGTTTATTGGAGGAGCTTTTAGATTATGGTGAGAGAT
 CTAACCAAGACAATTGTAACGACGCAACCGCTAGATCATCACTACTTACTTACAGTGTA
 ATGTTTTTGGAGTAAAGAGTAATAATCAATATAATACTTTAGTTTAGGAAAAA
 AAAAAA

>G1791 Amino Acid Sequence (domain in AA coordinates: TBD)

MERIESYNTNEMKYRGVRKRPWGKYAAEIRDSARHGARVWLGTFTNTAEDAARAYDRAAFG
 MRGQRAILNFPHEYQMMKDGPNGSHENAVASSSSGYRGGGGDDGREVIEFEYLLDSDLLE
 ELLDYGERSNQDNCNDNR*

>G1793 (59..1783)

AGTGATTTATTGATTAAACCAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
 GAATTCCTAACAACCTGGCTTGGCTTTCCTCTTTCACCGAACAACCTCTTCTTGCCTCCTCA
 TGAATACAACCTTGGCTTGGTCAGCGACCATATGACAACCTTTTCAAACACAAGAGTG
 GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
 CGATTTTCTCGGTGTGAGCAACCGGACGAAAACCAATCCAACACCTAGTAGCTTACAA
 CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
 CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACACAGTAGCTATCATGAGCTTCAAGA

GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
TGTAGACAAAGCTTCACCATCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
CGTTGTTGAGACGGCCACGCCAAGACGTGCATTGGACACTTTTCGGACAACGAACCTCGAT
CTATCGTGGTGTCAACAAGACATCGATGGACTGGTCGATATGAGGCTCATCTATGGGATAA
TAGTTGTAGAAGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACTTAAGTACTGGGGTCCCTTC
AACTACTACTAATTTCCCATTAACAACTACGAGAAAGAAGTAGAGGAAATGAAGCACAT
GACGAGACAAGAGTTTCGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
TTCGATGTATCGAGGAGTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
CCGAGTCGCCGGAACAAAGACCTCTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCCATCGGAGGAGG
CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
CACCTCATCAAGACTTCAGCTTCAACCTTACCTCTAAGCATTCACAACCATTAGAGCC
TTTTCTATCTCTTTCAGAACAAATGACATCTCTCATTACAACAACAACATGCTCACGATTC
CTCCTCTTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
CAATTACTTGCAGCAACAGTCGAGCCAGAACCTCAGCAGCTCTACAATGCGTATCTTCA
TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTTCTTGGGAACACGGTATTGG
TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAAACAGATTA
CGATATGCCCTTCCAGTGATGGAACCGGAGGGTATAGTGGTTGGACCAGTGAGTCTGTTCA
GGGGTCAAACCTTGGTGGTGTTCCTACTATGTGGAATGAGTAAACAAGGATCTCTTTCTT
GCGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates:179-255, 281-349)

MNSNNWLGFPLSPNSSLPPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGDEGGEVVKV
ADFLGVSKPDENQSNHLVAYNDSYFHTNSLMPVSQSNVVAACDSNTPNNSSYHELQ
ESAHLQSLTSLMGTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTRGYEAHLWDNSCRREGQSRKGRQVYLGGYDKEDKAARSYDLAALKYWGP
STTTNFPITNYEKEVEEMKHMTRQEFVAAIRRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFFSTEEAAEAYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESRRKREAEMLGSSSFQYGGGSGTSGSGSTSSRLQLQPYPLSIQPLE
PFLSLQNNDISHYNNNNNAHDSSSFNHHSYIQTQLHLHQQTNNYLQQQSSQNSQQLYNAYL
HSNPALLHGLVSTSIVDNNNNGGSSGSYNTAAFLGNHIGIGSSSTVGSTEEFPTVKTD
YDMPSSDGTGGYSGWTSSESVQGSNPGGVFTMWNE*

>G1795 (27..422)

ACAAACACGCAAAAAGTCATTAATATATGGATCAAGGAGGTGCGAGGTGTCGGTGCCGAGC
ATGGAAAGTACCGGGGAGTTCCGAGACGACCTTGGGGAAAATATGCAGCAGAGATACGAG
ATTTCGAGGAAGCACGGTGAACGTGTGTGGCTTGGAAACGTTCCGATACGGCAGAGGAAGCGG
CTAGAGCCTATGACCAAGCTGCTTACTCCATGAGAGGCCAAGCAGCAATCCTTAACTTCC
CTCATGAGTATAACATGGGGAGTGGTGTCTCTTCTTCCACCGCCATGGCTGGATCTTCCCT
CCGCTCCGCTCCGCTTCTTCTTCTTCTTAGGCAAGTTTTGAATTTGAGTACTTGGATG
ATAGTGTTTTGGAGGAGCTCCTTGAGGAAGGAGAGAAACCTAACAAGGGCAAGAAGAAAT
GAGCGAGATATAATTATGATTATTTCTAA

>G1795 Amino Acid Sequence (domain in AA coordinates: 12-80)

MDQGGRGVGAEHGKYRGVRRRPWGKYAAEIRDSRKHGERVWLGTFTDAEEAARAYDQAAAY
SMRGQAAILNFPHEYNMGSGVSSSTAMAGSSSASASASSSRQVFEEFYLLDDSVLEELLE
EGEKPNGKKK* —

>G1800 (61..894)

CCATTATCATATCCTCTTCTTCTTCTTCACTATCAATCTTCTTCTTCCACTACAACACAA
ATGGAGAAATCATCCTCAATGAAACAATGGAAGAAGGGTCTGCTCGGGGTAAAGCGGT
CCACAAAACGCTCTTTGTCACTACCGTGGAGTCAGGCAAAGGACTTGGGGCAAATGGGTG
GCTGAGATCAGAGAGCCCAAGAAGAGGGCAAGACTTTGGCTTGGCTCTTTTCGCTACAGCT
GAAGAAGCAGCTATGGCTTATGATGAGGCTGCCTTGAAACTCTATGGGCACGACGCATAC
CTCAACTTACCTCATCTTCAAGGGAATACAAGACCTTCTCTGAGTAACTCTCAGAGGTTT
AAATGGGTACCTTCAAGGAAGTTTATATCTATGTTTCTTCTCATGTGGTATGCTAAACGTG
AATGCTCAGCCTAGTGTTCACATAATCCAGCAAAGACTAGAAGAACTCAAGAAAACCTGGA

>G1811 (93..827)
AAAGGAGCATTGGTATCTCAAACAATATTTGCCCTTTCTCTATCTCTCTCATCACTAT
TTGCCATCTCTTTCTCTCTCCCTCTCTTTCAAATGTCAATAAACCAATACTCAAGCGATT
TCCACTACCATTCTCTCATGTGGCAACAACAGCAGCAACAACAACAACACCAAAACGACG
TCGTGGAAGAAAAGAAGCTCTTTTCGAGAAACCCCTTAACCCCAAGTGACGTGCGGAAAC
TCAACCGCTCGTCATCCCCAAACAGCAGCCGAGAGATAGTATCTCCCACTAGCGGCCCGC
CCGCAAGCGCCGTGGAGAAAGGACTTCTCTCTGCTTGTGAGCAGGAGGAAGTTAAACCAT

GGAGATTCAGATACTCGTACTGGAACAGTAGCCAGAGTTATGTCTTGACCAAAGGCTGGA
GCAGATACGTCAAGGAGAAGCACCTTGACGCCGAGAGCTCGTTCTCTCCATCGACACC
GTTGAGACGGCGGAAGATTCTTCATTGGCTGGAGAAGACGCGGTGACTCTTCTCTCCTCCT
CCGACTCTTATCGCCATGTTCAATCCAATGCCTCGCTCCAATATTATCCTCATGCAGGGG
CTCAAGCGGTGGAGAGCCAAAGAGGCAACTCGAAGACATTAAGACTGTTCCGAGTGAACA
TGGAGTGCCAGCTAGATTCCGACTGGTCCGAGCCATCCACACCTGACGGTTCACACAT
ATACAACCAATCAGCAGCAGTTTCTATTCTACCCCTCAACAACAACACTATCCTCCTCCGT
ACTACATGGACATAAGTTTACAGGAGATATGAACCGGACGAGCTAGAAGCCCACAAGGA
TTAAAAAAAAGCTTCACATCTGGTCTGTATGTTGTCATAGATGTTGATTCTCTTAATTT
TACACAAGCTTCATTTTGCATTATTTAAAGTAAATCGTATTTTGATTCTTCTTAAATC
TCTCTCAATTTTCACTCTCTCTCTTTTCTTCTTATGTATTAGATTCTTTTACATAGCTA
ACACTTGTATAGAGAATTCAAAGTCTGGCTATTTTCGAAAGTTATCTTTCTCTTAAAA
AAAAAA

>G1811 Amino Acid Sequence (domain in AA coordinates: TBD)
MSINQYSSDPHYHSLMWQQQQQQHQNVDVVEEKEALFEKPLTPSDVGKLNRLVLPKQHA
ERYFPLAAAAADAVEKGLLLCFEDEEGKPWRFRYSYWNSSQSYVLTKGWSRYVKEKHLDA
GDVVLFHRHRSDGGRFFIGWRRRGDSSSSSDSYRHVQSNASLQYYPHAGAAVESQRGNS
KTLRLFGVNMECQLDSDWSEPSTPDGSNTYTTNHDQHFYPPQQHYPPPYMDISFTGDM
NRTS*

>G182 (74..1366)

CGTCGACGATCAGATTCTTGGCTATAGCTGTATATATACACCAAGATACACTCATCATCG
TCATATATAGATTATGTCAGCGTCTCTGAGCTTCTTGACATGGAAAACCTCCAAGGAGA
CTTAACCGACGTCGTACGAGGAATCGGAGGCCACGTGTTATCACCGGAGACTCCTCCCTC
GAACATCTGGCCTCTTCTCTGTGCATCCAAACACCATCACCCTCAGATCTTAACATAAA
CCCCTTCGGAGATCCCTTTGTGAGCATGGACGATCCACTCCTCCAAGAACTAAACTCCAT
CACAACTCCGGCTATTTCTCACCGTAGGAGATAACAACAACAACATTCAACAACA
TGGTTTCTTGGTTCCAAAGGTATTTGAGGAGGATCATATAAAGAGTCAATGTAGTATCTT
CCCAAGAAATCCGGATCTCGCATAGTAACATCATCCACGATTCTTCTCCGTGTAATTCTCC
GGCCATGTCCGGCTCACGTTGTGCGAGCCGACGAGCCGCTCGCCGAGAGGCATCATCAA
CGTAGACACAAACAGTCTTAGAACTGTCTATTGGTTGATGGTACCACGTTCTCCTCGCA
GATTGAGATATCTTCCCTCGGAATCTAGGCCTTAAAGAAGGAAGAGTCAGGCAAAGAA
GGTGGTGTGATTCCGGCCCCGGCTGCAATGAACAGCCGATCAAGCGGAGAAGTGGTTCC
ATCGGATCTATGGGCTTGGCGTAAATACGGTCAAAAACCTATCAAAGGCTCTCCTTTTCC
AAGGGGTATTATAGATGCAGCAGCTCAAAAGGTTGTTGAGCAAGAAAGCAAGTCGAAAG
AAGCCGAACCGATCCAAACATGTTGGTGATTACATATACCTCCGAACATAACCATCCTTG
GCCCCATCAACGCAACGCTCTCGCCGGCTCCACAGCTCCTCCACCTCCTCCTCATCTAA
CCCTAATCCTTCCAAACCTCAACCGCAACGTAAACTCCTCATCCATTGGCTCCCAAAA
CACCATCTACTTGCTTCTCTCCACACTCCTCCTCTACCTCTCATCCTCCGCCATCAA
AGATGAACGAGGGGACGATTGGAGTTGGAAAACGTAGATGATGATGATAACCAGAT
TGCTCCATACAGACCGAGCTTCTATGATCATCAGCACCACCAACAGATGATTTCTTTGCAGA
TCTTGAAGAGCTAGAAGGAGATTCTCTAAGCATGTTGCTTTCTCATGGCTGTGGCGGCGA
CGGAAGGATAAAACGACCGCGTCCGATGGGATCAGCAATTTCTTCGGGTGGTTCGGGAGA
TAATAATTATAATAATTACGACGACCAAGACTCAAGGTCGTTATAGTATAGTGTTAATTA
CAGGTAAACAAATTATATTAAATTAAGTTGAGCTTGTGAAAATGAAGATCATATGGTCTG
GTCAGGTTGGGGG

>G182 Amino Acid Sequence (conserved domain in AA coordinates:217-276)
MCSVSELLDMENFQGLDLDVVRGIGGHVLSPETPPSNIWPLPLSHPTSPSDLNINPFGD
PFVSMDDPLLQELNSITNSGYFSTVGDNNNNIHNNGFLVPKVFEEDHIKSQCSIFPRIR
ISHSNIHDSPPCNSPAMSAHVVAASPRGIINVDINSPRNCLLDGTTFFSSQIQIS
SPRNLGLKRRKSQAKKVVCIPAPAMNSRSSGEVVPDLWAWRKYGQKPIKGSPPFRGY
RCSSSKGCSARKQVERSRDTPNMLVITYTSEHNHPWPIQRNALAGSTRSSSSSNPNPS
KPSTANVNSSSIGSQNTIYLPSTTPPPTLSSSAIKDERGDDMELENVDDDDNQAIPYR
PELHDHQHPDDFFADLEELEGLSMLLSHGCGDGDKTTASDGISNFFGWSGDNNYN
NYDDQDSRSL*

>G1835 (1..969)

ATGATTGGAACAAGCTTCCCCGAGGATCTTGATTGTGGCAACTTCTTTGACAACATGGAT
GATCTCATGGACTTTCCCGGTGGAGATATCGATGTCGGTTTCGGCATAGGTGACTCCGAC

TCTTTCCCTACCATCTGGACCACTCATCACGACACGTGGCCTGCCGCTTCTGATCCTCTC
TTCTCTTCCAACCAACTCTGATTCATCACCTGAGCTCTATGTTCCGTTTGAGGACATT
GTTAAGGTGGAAAGACCTCCAAGCTTTGTAGAGGAAACATTGGTTGAGAAGAAGGAAGAT
TCGTTTTTCGACAAACACTGATTCATCATCTTCTCATAGCCAATTCAGGAGCTCAAGTCCA
GTGTCGGTTCTCGAGAGCAGCTCCTCCTCGTCTCAAACCACCAACACAACCTCCCTTGTT
CTCCCTGGAAAGCAGGTCGTCACGCACAAAACGCCCTCGTCCACCTGTCCAGGATAAA
GATAGAGTCAAAGACAATGTGTGCGGTGGTGACTCGCGCTCATCATTAGAATACCGAAA
CAGTTTCTCTCTGATCACAACAAGATGATCAACAAGAAGAAGAAGAAGAGGCCAAGATT
ACTTCTTCTCTCTTCTCGTCCGGGATTGATCTTGAAGTCAATGGAACAACGTCGATTCTG
TATTCTTCAGAGCAATATCCGCTTAGGAAATGTATGCACTGTGAGGTACCAAGACTCCA
CAGTGGAGGCTTGGTCCAATGGGTCCAAAGACACTTTGCAATGCGTGGGTGTACGTTAC
AAATCAGGGAGGCTTTTCCCGGAGTACCGTCCAGCTGCTAGTCCAACATTTACTCCAGCT
CTTCACTCAAACCTCACACAAGAAAGTGGCTGAAATGAGAAACAAGAGATGCAGTGATGGT
AGCTACATAACCGAAGAGAATGATCTGCAAGGGCTGATTCGAACAATGCCTACATTGGC
GTAGACTAA

>G1835 Amino Acid Sequence (domain in AA coordinates: 224-296)
MIGTSFPEDLDCGNFFDNMDDLDFPGGDIDVGFIGSDSFPTIWTTHHDTWPAASDPL
FSSNTNSDSSPELYVPFEDIVKVERPPSFVEETLVEKKEDSFSTNTDSSSSHSQFRSSSP
VSVLESSSSSSQTNTTSLVLPKGHRPRTKRPRPPVQDKDRVKNVCGGDSRLIIRIPK
QFLSDHNKMINKKKKKKAKITSSSSSSGIDLEVNGNNVDSYSSEQYPLRKMHCETVTKTP
QWRLGPMGPKTLCNACGVRYKSGRLFPEYRPAASPTFTPALHSNSHKVAEMRNKRCSDG
SYITEENDLQGLIPNNAYIGVD*

>G1836 (47..610)
ATAACAAGCCTAGAACACTAGAAACTTCAAAAAAGAAAAAATCTTATGGAGAACAACAA
CGGCAACAACCACTGCCACCGAAAGGTAACGAGCAACTGAAGAGTTTCTGGTCAAAAGA
GATGGAAGGTAACTTAGATTTCAAAAATCACGACCTTCTATAACTCGTATCAAGAAGAT
TATGAAGTATGATCCGGATGTGACTATGATAGCTAGTGAGGCTCCAATCCTCCTCTCGAA
AGCATGTGAGATGTTTATCATGGATCTCACGATGCGTTCGTGGCTCCATGCTCAGGAAAG
CAAACGAGTCACGCTACAGAAATCTAATGTGATGCCCGAGTGGCTCAAACCTGTTATCTT
TGATTTCTTGCTTGATGATGACATTGAGGTAAAGAGAGAGTCTGTGTCGCCGCTGCTGA
TCCTGTGGCCATGCCACCTATTGACGATGGAGAGCTGCCTCCAGGAATGGTAATTGGAAC
TCCTGTTTGTGTTGATCTTGGAAATCCCAACCACAACCAAAATGCAGGCATGGCCTGG
AGCTTGGACCTCGGTGTCTGGTGAGGAGGAAGAAGCGCTGGGAAAAAGGAGGTGACGA
CGGAAACTAATAAGTGAATACGTTTTAGGGTATTTTCAAGGGAATATGTAGTAAATAGT
CATGGATC

>G1836 Amino Acid Sequence (domain in AA coordinates: 30-164)
MENNNNGNNQLPPKQNEQLKSFWSKEMEGNLDKFNHDLPIRIKKIMKYDPDVTMIASEAP
ILLSKACEMFIMDLTMRSLWHAQESKRVTLOKSNVDAVAQTVIDFLDDDDIEVKRESV
AAAADPVAMPPIDDGELPPGMVIGTFVCCSLGIHQPPQMQAWPGAWTSVSGEEEEARGK
KGGDDGN*

>G1838 (132..1628)
TTCCTTGGCATTCTCTTTAGAACTTTCGTACAAAATGCAAAACCTGAACCTCTAAAGCTA
AAAAAAAAGATTAGAGACTGTAAGTCTTTTATCAGATTTTCACTAGGAAAAAAGTTAC
AATCTTTTTTGTAGGCTCCTCCAATGACGAATTGCTTAACGTTTTCTCTGTACCAATGG
AGATGTTGAAATCAACTGATCAGTCTCACTTCTCTTCTTCTTACGACGATTCTTCTACTC
CTTATCTCATCGATAACTTCTATGCTTTCAAAGAAGAAGCTGAGATAGAAGCTGCTGCTG
CTTCAATGGCGGATTCAACAACCTTATCTACTTTTTTCGATCATTTCTCAGACTCAGATT
CAAAGCTGGAAGATTTCTCGGTGATTCTTTGTCCGTTACTCTGATAACCAACAGAGA
CCCAAGACTCTTCTTCTCTCTCACTCATTCTACGATCCACGTCACCGCACCCTTGCCGAAG
GAGTTACAGGGTTCTTCTCTGATCATCATCAGCCAGATTTCAAGACGATAAACTCGGGAC
CAGAAATCTTCGATGACTCAACAACCTTCCAACATCGGTGGTACTCATCTCTCCAGTCACG
TGGTGGAGTCATCAACGACGCGGAAGTTAGGGTTTAAACGGTGATTGCACCACCACCGGAG
GAGTTTTGTCTCTAGGGGTTAAACAACATCAGATCAACCTTTGAGCTGTAACAATGGCG
AGAGAGGTGGAAACAGTAACAAGAAGAAAACAGTTTCTAAGAAGGAAACATCAGATGATT
CAAAGAAGAAGATTGTGGAACATTGGGACAAAGAACTTCAATTTATCGTGGAGTCACCC
GACATAGATGGACTGGAAGATACGAAGCGCATCTATGGGATAACAGCTGTAGGAGGGAAG
GTCAAGCCAGAAAAGGACGTCAAGTGTAAGTTAGGTGGATATGACAAGGAAGATAGAGCAG

CTAGAGCCTATGACTTGGCAGCTTTAAATACTGGGGTTCTACTGCTACTACAAATTTTC
 CGGTCTCGAGTTATTCAAAGAAGTTGAGGAAATGAATCACATGACCAAGCAAGAGTTTA
 TTGCATCTCTTAGGAGGAAAAGTAGCGGTTTTTCGAGAGGAGCTTCAATATATAGAGGTG
 TCACAAGGCATCATCAACAAGGTCGCTGGCAAGCAAGAATCGGCCGTGTCGAGGAAACA
 AAGATCTTTACCTCGGAACCTTTGCAACCGAAGAGGAAGCAGCAGAGGCTTATGACATTG
 CAGCCATAAAGTTTCAAGGAATCAACGCAGTAACCTTTGAGATGAACAGGTATGACA
 TTGAAGCTGTCTGAATAGTTCTTTACCTGTAGGAGGAGCAGCTGCGAAACGCCACAAAC
 TCAAACTCGCTCTTGAATCTCCTTCTCATCATCTCTGACCATAACCTCCAACAACAAC
 AGTTGCTTCCGCTCCTCTTCTCCCTCGGATCAAAACCTAACTCAATCCCATGTGGCATT
 CATTGAGCCTTCAGTTCTCTATTACCACCAGAATCTTTTCAGCATTATCCTTTGGTCT
 CTGACTCTACAATTCAAGCTCCTATGAACCAAGCTGAGTTTTTCTTGTGGCCTAACCACT
 CTTACTAAATCATTGTTTGGTTCGTTCTTGCTTAGACTTCTATTACCGCCTAACCGATGAC
 CCGAGGCTTATCTTCTTGATTCTGGCTATAAGGATGAATCTTTCAAGTTCTTTTTTAAC
 TGTAGGCTTAAGACAGAAGTAGAGGGGAGAAAAGTTGAAGAATCTGAACTTTTGGGGTCA
 ATTTTGTATTAATGTTTTTCTTTTGTCAAGGGTGGATTATCGGTTTTTACTTATTTTT
 TGAATGTAATCGGCCTATAACGGTATAACTCTGTTTCCATTATGAATATTTTTCTCAA
 TTGAAAAA

>G1838 Amino Acid Sequence (conserved domain in AA coordinates: 229-305, 330-400)

MAPPMTNCLTFLSPMEMLKSTDQSHFSSSYDDSTPYLIDNFYAFKEEAEIEAAAASMA
 DSTTLLSTFFDHSQTQIPKLEDFLGDSFVRYSDNQTTQDSSSLTPFYDPRHRTVAEGVTG
 FFSDDHQPDKFTINSQPEIFDDSTTSNIGGTHLSSHVESSTAKLGFNGDCTTTGGVLS
 LGVNNSTSDQPLSCNNGERGNSNKKKTVSKKETSDDSKKIVETLGQRTSIYRGVTRHRW
 TGRYEHLWDNSCRREGQARKGRQVYLGGYDKEDRAARAYDLAALKYWGSTATTNFPVSS
 YSKELEEMNHMTKQEFIASLRRKSSGFSRGASIRGVTRHHQQGRWQARIGRVAGNKDLY
 LGTFATEEEEAAYDIAAIKFRGINAVTNFEMNRYDIEAVMNSSLPVGGAAAKRHLKLA
 LESPSSSSDHNLQOQQLPSSSPSDQNPNSIPCGIPFEPVLYYHQNFFQHYPLVSDST
 IQAPMNQAEFFLWPNQSY*

>G1843 (51..653)

CAGACATCACAATCAAATTAGGTCAGAAGATTAGTCGGAGAAAACAGCCATGGGAAGAA
 GAAAAGTAGAGATCAAACGAATTGAGAACAAAAGCTCTCGACAGTTACTTTCTGTAAAC
 GACGAAATGGTCTCATGGAGAAAGCTCGTCAACTCTCAATCTTTGTGAATCCTCCGTCG
 CTCTTATCATCATCTCTGCCACCGGAAGACTCTACAGCTTCTCCTCAGGTGATAGCATGG
 CCAAGATCCTCAGTCGTTATGAATTAGAACAGGCTGATGATCTTAAACCTTGGATCTAG
 AAGAAAAAAGCTTTAATTATCTTTTCGCACAAGGAGTTGCTAGAAACAATCCAATGCAAGA
 TTGAAGAAGCGAAAAGCGATAATGTAAGTATAGATTGTCTAAAGTCCCTGGAAGAGCAGC
 TCAAGACTGCTCTGTCTGTAAGTAGAGCTAGGAAGACAGAACTAATGATGGAGCTTGTGA
 AGACCCATCAAGAGAAGGAGAAGCTGCTGAGAGAGGAGAACCAGAGTTTGACTAACCCAGC
 TTATAAGATGGGGAAGATGAAGAAGTCTGTGGAAGCAGAGGATGCAAGAGCAATGTCAC
 CGGAAAAGTAGCTCTGACAACAAGCCACCGGAGACTCTCCTGCTTCTCAAGTAACCACCAT
 CACCAACGACTGATTCGAAAAATAAAATTGTAATAATTATGATTGTAGTTTCATAAGGA
 AAGCTACATACTGTATGTTAAAAATCCTCTTCTTCCCCCTGCTACGAAAAGTCATCCAA
 GGAGATGCATCAAATAAAGTAATTGATTTTTTATTGTTA

>G1843 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKVEIKRIENKSSRQVTFCKRRNGLMEKARQLSILCESSVALIIISATGRLYSFSSG
 DSMAKILSRYLEQADDLKTLDLEKTLNLYLSHKELLETIQCKIEAKSDNVSIDCLKSL
 EEQLKTALSVTARKTELMELVKTHQEKLLREBNQSLTNQLIKMGKMKKSVEAEDAR
 AMSPESSSDNKPPELTLKLLK*

>G1853 (1..1860)

ATGAGAGGTTCTTGGTACAAGAGTGTTTCTCTGTTTTTGGTCTCAGACCACGGATCAGA
 GGGTTGTATTCTTCATTGTTGGTGTGTGGCTCTAGTTACTATTTTAGCACCATTGACA
 TCTAATTCGTATGATTCCTTCGTCAGGTTGACACTGTGCCGAACATTTATAGTAAGTAT
 AGGAGGATAAAGGAGCAAGCTGCTGTTGATTATCTTGATCTGAGGTCTCTTCTTTAGGG
 GCTAGTTTAAAAGAGTTTCTTTTGTGGTAAAGAAAGAGAAAGTTATGTGCCTTGTTAT
 AACATAACTGGGAATTTGCTTGTGGGCTTCAAGAGGGTGAGGAGTTAGATCGACATTGC
 GAGTTTGAAAGAGAGAAGGAAAGATGTGTAGTTCTCTCCGAGAGATTAAAAATACCA
 CTTAGGTGGCCACTTGGTAGAGATATCATATGGAGTGGGAACGTGAAGATTACCAAAGAC
 CAGTTTCTTCTTCAGGAACGTGACAACGAGGTTAATGTTGCTTGAAGAGAATCAAATA

ACCTTTCACTCGGAGGACGGCCTGGTCTTTGATGGGGTCAAAGACTATGCTCGTCAAATT
GCTGAGATGATAGGTTTAGGAAGTGATACTGAATTTGCTCAAGCGGGTGACGGACTGTG
TTAGACATTGGTTGCGGATTTGGTAGCTTTGGTGCTCATTTAGTGTCTTTGAAGCTGATG
CCTATATGTATTGCTGAGTATGAGGCAACTGGGAGCCAAGTTCAGTTAGCTCTAGAGAGA
GGCCTTCTGCAATGATTGGCAATTTCTTTCAAACAGCTTCCTTATCCAGCACTGTCT
TTTGATATGGTCCATTGTGCTCAATGTGGCACTACTTGGGATATCAAAGATGCAATGCTA
CTTTTGGAAAGTGGATCGTGTCTGAAACCCGGGGGATACTTTGTTTTAACTTCTCCACA
AACAAAGCACAGGGAAACTTACCAGATACCAAGAAAACGAGCATCTCAACACGGGTGAAT
GAGTTATCTAAGAAAATCTGTTGGAGTCTAACAGCTCAGCAGGATGAGACGTTTCTTTGG
CAGAAAACCTCTGATTCAAGTTGCTATTCTTCTCGTTCGCAAGCTTCTATACCTCTTTGC
AAAGATGGAGATAGCGTTCCGTATTACCACCCATTGGTTCCATGTATAAGCGGAACCACG
AGTAAACGCTGGATTTCTATACAGAACAGGTCTGCTGTTGCAGGAACAACCTCTGCCGGG
CTTGAAATTCATGGTTTAAACCCGGAAGAATTCTTCGAGGATACACAAATATGGAGATCA
GCTCTGAAAACTATTGGTCTTTGCTTACACCTCTAATTTTCTCTGACCATCCGAAGAGA
CCCGGTGATGAGGATCCTCTCCCGCTTTCAACATGATACGCAATGTGATGGACATGCAT
GCTCGTTTGGGAATTTAAATGCCGCTTTACTCGACGAAGGAAAATCTGCTTGGGTAATG
AACGTCGTCACGTCATGCACGTAATACTCTTCTATCATACTTGATCGTGGTTTTCGCC
GGTGTCTACATGACTGGTGTGAACCATTCCCGACATATCCTCGAACATATGACATGCTT
CATGCCAATGAATTTCTCACACATCTTAGCTCAGAACGATGCAGCTAATGGACTTGTTC
TTGGAGATGGACCGGATTTCTTCGCCCTGAGGGATGGGTTGTTCTAAGCGACAAAGTGGGA
GTAATCGAGATGGCTCGAGCACTTGCAGCTCGAGTGCCTGGGAAGCAAGAGTCATTGAT
CTTCAAGATGGTAGTGACCAAAGACTTCTCGTCTGTCAAAAACCATTCATCAAAAAATAA
>G1853 Amino Acid Sequence (domain in AA coordinates: entire protein)
MRGSWYKSVSSVFLRPRIRGLLFFIVGVVALVTILAPLTSNSYDSSSSSTLVPNIYSNY
RRIKEQAAVDYLDLRLSLGASLKEFPFCGKERESYVPCYNITGNLLAGLQEGEELDRHC
EFEREKERCVRPRDYKIPLRWPLGRDIIWSGNVKITKDQFLSSGTVTTRLMLLEENQI
TFHSEDGLVFDGVKDYARQIAEMIGLSDEFAQAGVRTVLDIGCGFGSFGAHLVSLKLM
PICIAEYEATGSQVQLALERGLPAMIGNFFSKQLPYPALSFDMVHCAQCGTTWDIKDAML
LLEVDRLVKPGGYFVLTSPTNKAQGNLPDTKKTSTISTRVNELSKKICWSLTAQQDETFLW
QKTSDDSSCYSSRSQASIPCLKDGDSPYHPLVPCISGTTSKRWISIQNRSAVAGTTSAG
LEIHGLKPEEFFEDTQIWRSAKNYWSLLTPLIFSDHPKRPGEDEPLPPFNMIRNVMDMH
ARFGNLNAALLDEGKSAAVMNVVPNARNLPIILDRGFAGVLHDWCEPPTYPRTYDML
HANELLTHLSSERCSLMDLFLFEMDRILRPEGWVVLSDKVGVIEMARALAAVRWEARVID
LQDGSQRLLVLCQKPFIKK*
>G1855 (1..1902)
ATGGCGAAAGAGAACAGTGGTCATCATCACCAAACAGAAGCAAGAAGAAAGAACTAACT
TTGATTCTTGGTGTAAGTGGACTCTGCATTTTGTCTATGTTTATAGGTGCATGGCAAGCC
AATACCGTCCCATCTTCTATCTCGAAGCTCGGATGCGAGACGCAATCAAACCTTCTTCG
TCCTCTTCTCTTCTCATCTTCAGAGTCAGCTGAAGTATTTCAAAGCCATAATCAG
ATTGAGTTAAAGGAAACAAACCAACCATTAAGTACTTTGAACCATGTGAATTATCTCTC
AGTGAGTACACTCCTTGTGAAGACCGACAAAGAGGAAGAAGATTTCGATAGGAACATGATG
AAATATAGAGAAAGACATTGTCTGTAAAAGATGAGCTTCTTTATTGTTGATTCTCTCT
CCACCAAACACTACAAGATTCCATTTAAATGGCCACAAAGTAGAGACTATGCTTGGTATGAC
AATATCCCTCACAAGGAACCTTAGTGTGAGAAAGCAGTTCAAACTGGATTCAAGTTGAA
GGTGACCGCTTTAGATTCCCTGGTGGTACTATGTTTCTCGTGGAGCTGATGCTTAT
ATCGATGATATTGCTAGGCTTATTCTCTTACTGATGGTGAATCAGAACAGCTATTGAC
ACTGGATGTGGTGTGCAAGTTTGGTGCTTACCTCTTGAAGAGAGACATTATGGCTGTG
TCTTTTGCTCCAAGAGACACTCATGAAGCTCAGGTACAGTTTGCTTTAGAACGCGGAGTT
CCTGCGATAATCGGATATGGGATCAAGAAGACTTCTTATCCAGCTAGAGCTTTTGAT
CTTGCTCATTTCTCGTTGTTTGTATCCCTTGGTTTAAATGATGGTTTGTACCTTATG
GAGGTCGACCGGTTTAAAGACCGGGCGGTTACTGGATCTCTCGGGACCACCGATTAAC
TGGAAACAGTACTGGAGAGGGTGGGAGAGAACAGAGGAGGATTTGAAGAAAGAGCAAGAT
TCAATAGAAGATGTAGCAAAGAGTCTTGTGGAAGAAAGTAAGTGAAGAAAGGTAAGT
TCAATTTGGCAAAGCCTCTCAATCACATTGAGTGTAAGAAAGCTCAAACAAAACAATAAG
TCACCTCCGATATGCAGCTCAGATAACCGGATTCCGCTTGGTACAAAGACTTGGAAACT
TGTATAACACCATTACCAGAAACAAACAATCCAGATGATTGAGCAGGCGGTGCACTCGAG
GATTGGCCAGACCGGATTCGCGGTACCTCCAAGAATCATCAGAGGAAGTATACCAGAA

GATCTTAACTCCAATCCTCAAAGAGTACACGAGCCTGGATCACATAACCAAGAGGAGGGT
ATTGATGTAAATAACGCAACAATCACGAGCATGAATAG
>G1881 Amino Acid Sequence (domain in AA coordinates:5-28, 56-79)
MRILCDACESAAAI VFCAADEAALCCSCDEKVHKCNKLASRHLRVGLADPSNAPSCDICE
NAPAFFYCEIDGSSLCLQCDMVVHVGGKRTTHRRFLLLRQRIEFPDGKPNHADQLGLRCQK
ASSGRGQESNGDHDHNMIDLNSNPQRVHEPGSHNQEEGIDVNNANNHEHE*
>G1882 (1..1200)
ATGGTTTTTCTTCATTTCTTACTTATCCTGATCATTTCATCAAACCTGGCAACAACAACAT
CAACCAATCACAACCACCGTTGGATTACGGGAAATAACATCAACCAACAGTTTCTTCCT
CACCATCCCCTCCCACCGCAACAGCAACAACGCCTCCGCAGCTTCACCACAACAACGGT
AACGGCGGAGTCGTGTTCCTGGTGGACCTGGCGGGTTAATCCGACCAGGTTTCGATGGCG
GAAAGAGCAAGGCTAGCCAACATACCTTACCTGAAACAGCCTTGAAGTGTCCAAGATGT
GACTCAACTAACACCAAAATCTGTCTTCAACAACACTACAGTCTCACTCAACCTCGCCAC
TTCTGCAAAGCATGCCGTGTTACTGGACACGTGGCGGTGCTCTAAGGAGCGTTCCCGTC
GGTGGCGGTTGCCGTAGAAACAAAAGAACCAAAACAGCAGCGGTGGAGGTGGCGGTAGC
ACCAGTAGCGGTAACAGCAAGTCACAAGACAGCGCCACGAGCAACGACCAATACCACCAC
CGAGCCATGGCTAACAAATCAGATGGGACCACCTTCTTCGTCTATCGTCTTAAGCTCGTTG
CTGTCTTCTTACAACGCGAGGGTTAATCCCCGACATGATCATAACAGCAATAACAACAAC
ATACTTGGACTTGGATCATCTTTGCCTCCTCTTAAGCTTATGCCTCCTTAGACTTCACA
GACAACTTCACCTTACATACGGTGCCGTTTCAGCTCCTTCTTATCATATAGGCGGTGGA
AGCAGTGGAGGAGCGCGGCTCTTTTAAACGGTTTTTGACCAGTGGAGATTCCCGGCAACA
AACCAACTTCTTTTAGCGGTTTAGACCCGTTTGATCAACAACATCAAATGGAGCAGCAG
AATCCAGTTTACGGATTGGTTACCGGGTCGGGTCAAGTATCGACCTAAGAACATTTTCCAT
AACCCTATCTCTCTTCTTCTGCTGCTTCATCAGCTATGGTTACAGCCACCGCGTCGCAA
TTAGCTTCAGTGAAAATGGAAGATAGTAACAATCAGCTCAACTTGTCTAGACAACTTTTT
GGAGACGAACAACAGCTCTGGAATATTATGCGCTGCTGCAGCATCCACCGCAGCTGCA
ACAAGTTCGTGGAGTGAAGTCTCTAATAATTCAGTTCTTCTTCTACTAGCAATATATAA
>G1882 Amino Acid Sequence (domain in AA coordinates:97-125)
MVFSSFTYPDPHSSNWQQHQPIITTVGFTGNNINQQFLPHHPLPPQQQTPPQLHHNNG
NGGVAVPPGGPGLIRPGSMAERARLANIPLPETALKCPRCDSTNTKFCYFNNSLTQPRH
FCKACRRYWRGGALRSVPVGGGCRNRKRTKNSSGGGGSTSSGNSKSQDSATSNQYHH
RAMANNQMGPPSSSSSLSSLLSSYNAGLIPGHDHNSNNNNILGLSSLPPLKLMPLDFT
DNFTLQYGAVSAPSYHIGGGSSGGAALLNGFDQWRFPATNQLPLGLDLPFDQHQHMEQQ
NPGYGLVTGSGQYRPKNIFHNLISSSSSASSAMVTATASQLASVKMEDSNNQLNLSRQLF
GDEQLWNHIGAAAATAAATSSWSEVSNNFSSSSTSNI*
>G1883 (1..1110)
ATGGACGCTACGAAGTGGACACAGGGTTTTCAAGAAATGATGAACGTTAAACCAATGGAG
CAGATCATGATTCTTAATAACAACACACATCAACCAAACACCACATCCAATGCAAGGCCA
AACACCAATTCTACATCTAACGGCGTCTCAACTGCTGGAGCAACCGTCTCCGGCGTAAGC
AACAAACAATAACAATACGGCGGTGTGGCGGAGAGGAAAGCAAGACCACAAGAGAAACTA
AATTGTCCAAGATGCAACTCAACCAACACAAAGTTTTGTTACTACAACAACATATAGTCTC
ACACAACCAAGATACTTCTGCAAAGGTTGTGCAAGGTATTGGACCGAAGGTGGATCTCTT
AGGAATGTTCTGTGGGAGGAAGCTCAAGAAAGAACAGAGATCATCTTCATCTTCTTCA
TCAAACATCCTTCAGACAATACCATCTTCACTTCAGATCTAAACCCGCCAATACTCTTC
TCAAACCAAATCCATAATAAATCGAAAGGGTCATCACAAGATCTCAACTTGTGTCTTTC
CCAGTCATGCAAGATCAACATCATCATGTCCATATGTCTCAGTTTCTTCAGATGCCT
AAGATGGAGGGGAAGGTGAACATAACTCATCAGCAGCAGCCTTCATCATCTTCTTCTGTC
TATGTTCTCTCGTCTCTCTGTTTCAGCTCTTGAACCTTTAAGAACCGGAGTTAATGTT
TCTTCAAGATCAGGGATTAACTCATCGTTTCATGCTTCCGGTTCAATGATGGATTCAAAC
ACTGTGCTTTACACTTCTTCAGGGTTTCCAACAATGGTGGATTACAAGCCAAGTAATCTC
TCCTTCTCTACCGATCATCAAGGGCTTGGACACAATAGCAACAATAGGTCTGAAGCTCTT
CATAGTGATCATCACCACAAGGTAGAGTTTTGTTTCCATTGTTGGGATCAAATGAAGGAG
CTTTTCATCAAGCATAACACAAGAGTTGATCATGATGATAATCAACAACAGAAGAGTCAT
GGAAATAATAATAATAATAAATCAAGCCCTAATAATGGATATTGGAGTGGGATGTTT
AGTACTACAGGAGGAGGATCTTCATGGTGA
>G1883 Amino Acid Sequence (domain in aa coordinates: 82-124)
MDATKWTQGFQEMMNVKPMEQIMI PNNNTHQPNTTSNARPNTILTSNGVSTAGATVSGVS

NNNNNTAVVAERKARPQEKLNCPRCNSTNTKFCYNNYSLTQPRYFCKGCRRYWTEGGSL
RNPVVGSSSRKNNRSSSSSSSSNILQTIPISSLPDLNPPILFSNQIHNKSKGSSQDLNLLSF
PVMQDQH HHVHMSQFLQMPKMEGNGNITHQQQPSSSSSSVYSSSSSPVSALELLRTGVNV
SSRSGINSSFMPSGSMDSNTVLVTSSGFPTMVDYKPSNLSFSTDHQLGHNSNNRSEAL
HSDHHQQGRVLFPPFDQMKESSSITQEVDDHNNQQQKSHGNNNNNNSSPNNGYWSGMF
STTGGGSSW*

>G1884 (1..741)

ATGATGACGTCAATCCATCAGAGCAACACCACCGGCTTTAAACCGCGGCGGATCAAGACG
ACGGCGAAGCCACCACGTCAATCAATAACAAAGAACCATCTCCGGCGACGACGCGGTG
CTCAAGTGTCCGAGATGTGATTCAATCAACACCAATCTGCTACTACAACAACACTACAGC
TTGTCTCAGCCACGTCACTACTGCAAGAACTGTCGTCGTTACTGGACACGTGGCGGCGCC
CTCCGTAACGTTCCCATCGGTGGCTCCACTCGAAACAAGAACAAGCCTTGACAGCTCCAA
GTCATCTCTTCTCTCTCTTGTCTCGAACGGGACGTCACTCGGCGTCTCGTGAGCTTGTA
AGAAACCATCCATCGACGGCAATGATGATGAGTCTGCTGGTGGATTCTCCGGCTATATG
TTCCGTTGGATCCTAACTTCAACCTTGCCTCGTCTTCTATCGAGTCTTTGAGTTCTTTT
AACCAAGATTTGCACCAAGCTTCAGCAACAAAGACTCGTCACTTCCATGTTTCTCCAA
GATTCTCTTCCGGTTAACGAGAAAACGGTTATGTTTCAGAACGTAGAGTTGATTCTCTCT
TCGACGGTGACGACGATTGGGTTTTCGATAGGTTGCCACTGGAGGAGGTGCAACAAGT
GGCAATCATGAAGATAATGATGATGGGGAGGGTAATTTGGGAAATTGGTTCCATAATGCT
ATAATAATGCTCTGCTCTAA

>G1884 Amino Acid Sequence (domain in AA coordinates: 43-71)

MMTSSHQSNNTGFKPRRIKTTAKPPRQINNKEPSPATQPVLCPCRDVNTKFCYNNYS
LSQPRHYCKNCRRYWTTRGGALRNVPIGGSTRNKNKPCSLQVISSPPLFSNGTSSASREL
RNHPSTAMMMSSGGFSGYMFPLDPNFNLAASSIESLSFNQDLHQLQQORLVTSMFLQ
DSLVPNEKTVMFQNVLEIPPSTVTTDWFVDFRATGGGATSGNHEDNDGEGNLGNWFHNA
NNNALL*

>G1891 (1..750)

ATGGATAACTTGAATGTTTTTCGCAATGAAGACAATCAAGTGAATGATGTGAAGCCCCCA
CCACCACCACCTCGAGTGTGTGCAAGGTGTGATTCTGATAATACTAAATTTTGTTATTAC
AACAACTACTGTGAGTTTCAAGCCACGATACTTCTGCAAGAACTGTCGTAGATACTGGACT
CATGGTGGGGCTTTAAGAAACATACCAATTGGTGGAGTAGTCGTGCCAAACGGGCAAGG
GTAAATCAACCTTCGGTTGCTCGGATGGTTTCTGTTGAGACCAACGAGGTAACAATCAA
CCTTCTCTAATGTTCAAGAAAACGTTTCATCTTGTGGATCTTTTGGTGCTTCATCTTCA
TCTTCTGTTGGTGCTGTTGGGAACCTTTTGGTTCTTGTATGATATTATGTTGGTGATG
GTAAACAAATTTGCATCAACTCGAAGTTCGACCAATCATCGCTTAGCTTTCCATGAT
GGATCATTTGAGCAAGACTATTACGATGTTGGGTCCGATAATCTTTGGTCAACCAACAA
GTTGGTGGCTACGGTTATCACATGAATCCAGTGGATCAATTCAAGTGGAAACAGAGCTTC
AACAACTATGAACATGAATTATAATAACGATAGCACTAGTGGAAGTAGCAGAGGATCT
GACATGAATGTGAACCATGATAACAAGAAGATCAGATACCGCAACTCTGTGATTATGCAT
CCTTGTCATCTGGAGAAGGATGGTCCTTGA

>G1891 Amino Acid Sequence (domain in aa coordinates: 27-69)

MDNLNVFANEDNQVNDVKPPPPPPRVCARCDSNTKFCYNNYCEFPQRYFCKNCRRYWT
HGGALRNPIIGSSRAKRARVNQPSVARMVSVETQRGNNQPFQENVHLVGSFGASSS
SSVGAVGNLFGSLYDIHGMVITNLHPTRTVRPNHRLAFHDGSFEQDYDVGSDNLLVNQQ
VGGYGYHMPVDQFKWNQSFNNTMNMNYNDSTSGSSRGSMDMNVNHDNKKIRYRNSVIMH
PCHLEKDG*

>G1896 (1..951)

ATGTCCTCCCATACCAATCTCCCTCTCCCAAACAGTTTCTAAACAGATCACCGTATC
TCCGGTACATCCCAAACCAAGAAACCACCGTCTTCTCCGTAGCTCAAGACCAACAAAC
CTAAATGCCCTCGTTGCAACTCTCCAAACACAAAGTTCTGTTACTACAACAACACTACAGT
CTCTCTCAACCTCGTCACTTCTGCAATCTGTGCGCGTTACTGGACACGTGGCGGTGCT
CTAAGAAACGTCCCATCGGTGGTGGTTGCCGGAAAACCAAAAATCTATCAAACCTAAT
TCCTCCATGAACACACTTCTTCTGCTCTTCTCTCTCAGAGGTTCTTCTCATCAATCATG
GAAGATTATCCAAATCTTCCCTCTCCGACAACAATGGATTTTCAGCTGGCCGGATTA
TCTCTCAACAAATGAACGATCTTCACTTTTGAATAACCAAGAGTTCTTGATCTTAGG
CCCATGATGTCCTCGGGCCGAGAAAACACACCCGTTGATGTCGGGTGGGTTTATCCCTA
ATGGGTTTTGGAGATTTCAACAACAACCATTCACCGACGGGTTCAACAACCGCCGGAGCA

AGCGACGGAACTTAGCTTCTTCTATAGAGACTTTGAGTTGTTTAAACCAAGATTACAC
TGGAGGCTTCAGCAACAGAGGATGGCGATGCTTTTTGGTAATTCTAAGGAAGAACTGTT
GTCGTCGAGAGGCCACAACCTATTCTTTATCGGAATCTTGAGATCGTAAACTCATCATCG
CCGTCGTCGCCGACGAAGAAAGGAGATAATCAGACAGAGTGGTATTTTGGTAATAACAGT
GATAATGAAGGAGTGATAGTAATAATGCTAATACAGGAGGAGGAGGAAGTGAATGGAAC
AATGGAATTCAAGCTTGGACTGATCTTAATCATTATAATGCATTGCCTTGA

>G1896 Amino Acid Sequence (domain in aa coordinates: 43-85)
MSSHTNLPSPKPVKPDHRISGTSQTKPPSSSSVAQDQQLKPCRCNSPNTKFCYNNYS
LSQPRHFCKSCRRYWTRGGALRNVPVIGGCRKTKKSIKPNSSMNTLPSSSSSQRFFSSIM
EDSSKFFPPPTTMDFQLAGLSLNKMNDLQLLNNQEVLDLRPMSSGRENTFVDVGSGLSL
MGFGDFNNHNSPTGFTTAGASDGNLASSIETLSCLNQDLHWRLQQORMAMLFGNSKEETV
VVERPQPILYRNLEIVNSSSPSPTKKGDNDQTEWYFGNNSDNEGVISNNANTGGGGSEWN
NGIQAWTDLNHYNALP*

>G1898 (1..630)
ATGCCGTCGGAACCAACCAACCCGACCCACCAGAGTTCAGCCCTCAACGCGCGCTTAC
CCACCGCCAAATCTGGCTGAGCCTCTTCCTTGTCCTCGCTGCAACTCCACCACCAAG
TTCTGTTACTACAACAATAACCTCGCTCAGCCTCGCTACTACTGCAATCTTGCCGC
CGTTACTGGACTCAAGGTGGTACACTCCGTCGCTCCCGTCGCTGGTGGAACTCGTCGA
AGCTCCTCAAAACGTCACCGTTCTTTCTCCACCACCTGCCACCTCCTCTCTCTCTTCT
TCCGTCATCACCACACGACACAAGAACCAGCCACGACTGAAGCGAGTCAAACCTAAGGTT
ACTAATTTAATTCAGGTCATGGAAGCTTTGCTTCTCTGTTAGGTTTAGGAAGTGGAAAT
GGTGGGTTGGATTACGGGTTTGGGTACGGGTACGGGCTTGAGGAGATGAGTATTGGGTAT
CTTGAGATTCTTCCGTAGGAGAGATTCCGGTGGTTGATGGTTGTGGTGGTGACACGTGG
CAGATTGGGGAGATTGAAGGTAAAGTGGAGGAGACAGTTTGATATGGCCTGGTCTTGAG
ATCTCAATGCAAACCAACGATGTTAAGTGA

>G1898 Amino Acid Sequence (domain in AA coordinates: 31-59)
MPSEPNTQTRPVRVQSTAAAYPPPDLAEPLPCPRCNSTTKFCYNNYNLAQPRYCKSCR
RYWTQGGTLRDVVPVGGGTRRSSSKRHSFSTTATSSSSSSSVITTTTQEPATTEASQTKV
TNLISGHGSFASLLGLSGNGGLDYGFYGYGLEMSIGYLGDSVGEIPVVDGCGGDTW
QIGEIEGKSGGDSLIWPGLEISMQTNDVK*

>G1902 (1..615)
ATGCAGGATCCAGCAGCATATTACCAGACGATGATGGCGAAGCAACAACAACAACA
CCACAGTTTGCAGAGCAAGAACAGTTAAAGTGTCCTCGTTGTGACTACCAAAACTAAA
TTCTGTTACTACAACAATACTCTCACAGCCTCGTCACTTTTGCAAAAGCTGTCTGT
CGTTACTGGACTAAAGGCGGCGCTCTCCGTAACGTTCCCGTCGGTGGTGGTCTCGTAAG
AACGCAACCAACGATCCACTTCTTCTTCTTCTTCTGCTTCTCTCTTCCAACAGTAGC
CAAAACAAGAAGACGAAAAACCCGGATCCGGATCCTGATCCACGTAATTTCTAAAAACCG
GATTTGGATCCGACCCGGATGCTTTACGGGTTTCCGATCGGTGACCAAGACGTGAAGGGT
ATGGAGATTGGTGAAGCTTTAGCTCGTTGTTGGCGAATAATATGCAGCTTGGTCTTGA
GGAGGAGGGATCATGCTTGACGGGTCGGGTTGGGATCATCCGGGTATGGGTTTGGGTTT
AGGAGAACCGAACCGGTAATAATAATAATAACCCATGGACCGATCTGGCTATGAACAGA
GCGGAGAAAACTGA

>G1902 Amino Acid Sequence (domain in AA coordinates: 31-59)
MQDPAAYYQTMMAKQQQQQQPQFAEQEQLKCPRCDSPTNTKFCYNNYNLSQPRHFCKSCR
RYWTKGGALRNVPVGGGSRKNATKRSTSSSSSSASSPSNSSQNKKTKNPDPPDPRNSQKP
DLDPTRMLYGFPIGDQDVKGMEIGGSFSSLLANNMQLGLGGGGIMLDGSGWDHPPMGLGL
RRTEPGNNNNPNWTDLAMNRAEKN*

>G1904 (1..924)
ATGCAAGATATTATGATTTCTCCATGAACGGAGTTGGTGGTGGGGGAGGAGGAGGG
AGGTTTTTTCGGTGGAGGAATCGGCGGCGGAGGAGGTGGTGATCGAAGGATGAGAGCTCAT
CAGAACAATATACTTAACCATCATCAATCTCTCAAGTGTCTCGTTGTAATTCTCTTAAC
ACAAAGTTCTGTTACTACAACAATTACAATCTTCTCAGCCTCGTCACTTTTGCAAGAAC
TGTCGTCGTTACTGGACTAAAGGTGGTGTCTCCGTAACGTTCCCGTCGGAGGTGGTTGC
CGGAAAGCTAAAGCTTCAAAAACAAACAGGTTCCGTCGTCGTCATCAGCCGACAAACCA
ACGACGACGCAAGATGATCATCAGTGAGGAGAAATCGAGTACAGGATCTCACTCTAGC
AGCGAGAGCTCTTCTCTCACCGCTTCTAACTCTACCACCGTCGCCCGCTCTCCGTCACC
GCGGCGGCGGAAGTTGCTTCGTCGGTTATTCAGGTTTTGATATGCCTAATATGAAAATT

TACGGTAACGGGATCGAGTGGTCGACGTTACTTGGACAAGGCTCATCGCCGGTGGTGT
TTCTCGGAGATCGGTGGTTTTCCGGCGGTTTCAGCTATTGAACTACACCGTTTGGATT
GGGGGTAAATTCGTAAATCAAGATGATCATCTGAAGTTAGAAGGTGAACTGTACAGCAG
CAACAGTTTGGAGATCGAACGGCTCAGGTTGAGTTTCAAGGAAGATCTTCGGATCCGAAT
ATGGGATTTGAACCGTTGGATTGGGGAAGTGGCGGTGGAGATCAAACACTGTTTGATT
ACCAGTACCGTTGATCATGCATACGAGTCAAAGTCAATGGACGTCGTCTGACCAAGAT
CAGAGTGGTCTCTACCTTCCTTGA

>G1904 Amino Acid Sequence (domain in aa coordinates: 53-95)
MQDIHDFSMNGVGGGGGGGRFFGGGIGGGGGGDRMRHQNILNHHQSLKCPRCNSLN
TKFCYNNYNLSQPRHFCKNCRRYWTKGGVLRNVPVGGGCRKAKRSKTKQVPSSSSADKP
TTTQDDHHVEEKSSSTGSHSSSESSSLTASNSTTVAAVSVTAAAEVASSVIPGFDMPNMKI
YGNGLIEWSTLLGQSSAGGVFSEIGGFPAVSAIETTFPGFGGKFVNQDDHLKLEGETVQQ
QQFGDRTAQVEFQGRSSDPNMGFEPLDWGSGGGDQTLFDLTSTVDHAYWSQSQWTSDDQD
QSGLYLP*

>G1906 (1..795)
ATGGTGGAACGTGCTCGGATCGCAAAAGTCCCATTGCCTGAAGCAGCTCTAAATTGCCCT
AGATGTGACTCAACCAATACTAAGTTCTGTTACTTCAATAACTATAGCCTTACTCAACCT
CGCCATTTCTGCAAAACATGTCGTCGCTATTGGACACGTGGCGGTTCTTGAGGAATGTT
CCTGTTGGAGGAGGCTTTAGGAGGAACAAGAGAAGCAAATCCAGATCGAAATCTACGGTC
GTGGTCTCGACTGATAATACTACTAGTACTTCATCACTTACTTCTCGCCCAAGTTACTCA
AACCCTAGCAAGTTTCATAGCTACGGTCAAATCCCGGAGTTTAAATCCAACCTTGCCCATC
TTGCCTCTCTCCAAAGCCTTGGAGATTACAATTCAAGCAACACTGGATTAGATTTTGGT
GGAACCTCAAATAAGCAACATGATAAGTGGTATGAGTTCTAGTGGTGGGATCTTGGATGCA
TGGAGAATACCTCCATCACAACAAGCTCAGCAATTCCTTTCTTGATCAACACTACCGGA
TTGGTGCAATCTTCAAACGCGTTATATCCATTACTAGAAGGCGGGGTTAGCGCCACGCAA
ACAAGAAATGTGAAGCGGAAGAGAATGATCAGGATCGGGGTAGGGATGGGGATGGAGTG
AATAACTTATCAAGAACTTTTTGGGTAATATCAACATAAACTCAGGCAGGAACGAGGAA
TACACATCATGGGGAGGTAACAGTTCTTGGACCGGTTTCACCTCCAACAACCTCAACAGGC
CATCTCTCATTCTAA

>G1906 Amino Acid Sequence (domain in AA coordinates: 19-47)
MVERARIAKVLPEAALNCPKRDSTNTKFCYFNYSLTQPRHFCKTCRRYWTRGGSRLNV
PVGGGFRNRKRSKRSKSTVVVSTDNTTSTSSLSRPSYSNPSKFHSYQIPEFNSNLPI
LPPLQSLGDNSSNTGLDFGGTQISNMISGMSSSGILDWRIPPSQQAQQFPFLINTTG
LVQSSNALYPLLEGGVSATQTRNVKAEENDQDRGRDGDGVNNLSRNFLGNININSGRNEE
YTSWGGNSSWTGFTSNNSTGHLSF*

>G1913 (1..744)
ATGGAGAGAGCAGAGGCCTTGACATCATCGTTTATATGGCGGCCAAACGCAAACGCAAAC
GCGGAGATCAGCCGAGTTGTCCAAGATGTGGATCCTCTAACACAAAGTTCTGTTACTAC
AACAACATATAGCCTCACTCAGCCTCGCTACTTCTGCAAAGGCTGCCGCAGATATTGGACC
AAAGTGTTTCCCTCCGCAATGTTCCTGTAGGCGGTGGCTGTGCGAAATCCCGCCGCCCC
AAATCATCTTCTGGTAACAATACTAAAACTAGCCTAACCGCTAATCTTGGCAACCCCGGT
GGTGGTTCACCAAGCATCGATCTTGCTCTTGTTTACGCCAATTTCTTGAATCCAAAGCCT
GACGAATCTATACTACAAGAAAATTGCGACTTAGCCACTACGGATTTTTTGGTAGATAAT
CCTACCGGCACTTCCATGGACCCTTCATGGAGTATGGACATCAATGATGGTCATCATGAT
CATTATATTAATCCGGTGGAAACATTTGTGGAGGAATGTGGTTATAATGGCTTGCCCTCCA
TTTCTGTGTAAGAGCTTCTCTTTAGACACTAATGGTGTGTTGGTCTGATGCTTTGTTG
ATTGGTCATAACCATGTAGACGTTGGCGTGAATCCCGTTTCAAGGCTGTACACGAACCGGTG
GTTTATTTCGCTGAAGAATCCAATGATTCCACCAATCTCTTGTGTTGGAAGTTGGAGCCCT
TTTGATTTCACTGCCGATGGATGA

>G1913 Amino Acid Sequence (domain in AA coordinates: 27-55)
MERAELTSSFIWRPNANANAEITPSCPRCGSSNTKFCYNNYSLTQPRYFCKGCRRYWT
KGGSLRNVVGGGCRKSRPKSSSGNNTKTSLTANSNPGGGSPSIDLALVYANFLNPKP
DESILQENCDLATDFLVDNPTGTSMDPSWSMDINDGHHHDHYINPVEHIVEECGYNGLPP
FPGEELSLDINGVWSDALLIGHNVHDVGVTVPQAVHEPVVHFADESNDSTNLLFGSWSP
FDFTADG*

>G1914 (1..945)
ATGGAGAGATACAAGTGTAGATTTTGCTTCAAGAGCTTCATCAATGGAAGAGCTTTAGGT

GGTCACATGAGATCTCACATGCTTACTCTTTCTGCAGAACGTTGTGTAATAACTGGTGAA
GCAGAAGAAGAAGTAGAGGAACGGCCGAGTCAACTCTGTGACGACGACGACGATACCGAG
TCCGATGCTTCTTCTTCTTCTGGTGAGTTTGATAATCAAAAGATGAATCGTCTTGATGAT
GAATTGGAGTTTGATTTCTGCTGAAGACGACGACGTTGAAAGTGAAACCGAGTCGTCCAGG
ATTAACCCAACTCGGCGACGATCTAAACGAACTCGGAACTTGGATCGTTTGATTTGAC
TTTGAGAAGCTAACAACGAGCCAACCCAGTGAGTTAGTGGCCGAGCCAGAGCATCACAGC
TCAGCTTCTGATACAACAACGAGGAAGATCTCGCCTTTTGTCTCATTATGCTGTCCAGA
GACAAATGGAAGCAACAGAAGAAGAAGCAACGTGTAGAAGAAGATGAGACAGATCAT
GACAGTGAAGATTACAAATCAAGCAAGAGCAGAGGGAGATTCAAGTGTGAGACTTGTGGT
AAAGTGTTTAAATCGTATCAAGCATTAGGAGGACACAGAGCAAGCCACAAGAAGAACAAG
GCATGCATGACGAAAACAGAGCAAGTTGAAACAGAGTACGTTCTTGAGTAAAGGAGAAG
AAAGTTCATGAATGTCCGATCTGTTTTAGGGTTTTTACTTCAGGGCAAGCATTGGAGGT
CATAAGAGATCTACGGAAGTAACATCGGAGCAGGAAGAGGATTGTGAGTAAAGTCAAATT
GTCCAAATCGAAGAAGAAGTATCAGTGAAACAGAGGATGATTGATCTTAATCTTCCTGCA
CCTAATGAAGAAGATGAAACTTCTTTGGTGTGTTGATGAATGGTGA

>G1914 Amino Acid Sequence (domain in AA coordinates:195-216, 245-266)

MERYKCRFCFKSFINGRALGGHMRSHMLTSLAERCVITGEAEVEEVEERPSQLCDDDDDE
SDASSSSGEFDNQMNRLDDELEFDFAEDDDVESETESSRINPTRRRSKRTRKLGSDFD
FEKLTTSPSELVAEPEHHSSASDTTTEEDLAFCLIMLSRDKWQKQKKKQKQVEEDTDH
DSEYDKSSKSRGRFKCETCGKVFKSYQALGGHRASHKKNKACMTKTEQVETEVVLGVKEK
KVHECPICFRVFTSGQALGGHKRSHGSGNIGAGRGLSVSQIVQIEEEVSVKQRMIDLNLPA
PNEEDETSLVFDEW*

>G1925 (1..945)

ATGGAAGAAAATCTTCTCCGGGGTTTCAGATTTTCATCTACAGACGAGGAGCTCATAACG
CATTATCTATGTCGGAAAGTCTCCGATATAGGATTACCGGTAAAGCTGTGTCGACGTT
GATCTCAACAAGTGTGAACCTTGGGATTTGCCAGCCAAGGCTTCAATGGGAGAGAAAGAG
TGGTATTTCTTCAGCCAAAGGGATCGGAAATATCCAACCGGTTAAGAACAACCGGGCA
ACAGAAGCTGGTTACTGGAAAACCAACCGGAAAGATAAAGAAATATACCGAAGTGGAGTG
TTGGTTGGGATGAAGAAAACCTAGTTTTCTACAAAGGAAGAGCTCCCAAAGGTGAGAAA
AGCAATTGGGTTATGATGAGTACAGGCTTGAGAGCAAACAACCTTTCAACCCACGAAT
AAGGAGGAATGGGTAGTGTGTAGGGTTTTCGAAAAGAGCACGGCAGCAAAGAAAGCACAA
GAACAACAACCTCAATCTTCTCAACCATCTTTGGATCTCCATGCGATGCAAACTCATCA
ATGGCAAATGAGTTTGAAGATATTGATGAGCTTCCGAATCTGAATTCAAACCTCATCAACC
ATCGATTACAATAATCATATCCATCAATATTCGCAACGCAATGTTTACTCAGAAGACAAC
ACAACAAGTACGGCTGGTCTCAACATGAACATGAACATGGCTAGTACTAATCTTCAGTCT
TGGACAACAAGTCTCCTTGGTCCGCCTTTATCTCCAATCAACTCTTTGTTGCTCAAGGCT
TTCCAAATCAGGAACCTTTATAGTTTCCCAAAGAGATGATCCCCAGTTTCAATCATTTCT
TCTCTTCAACAAGGAGTCTCCAATATGATCCAAAATGCTTCAAGTTCTGCTCTCAAGTGCAA
CCCCAACCGCAAGAGGAAGCGTTTAATATGGACTCCATATGGTGA

>G1925 Amino Acid Sequence (conserved domain in AA coordinates:6-150)

MEENLPPGFRFHPTDEELITHYLCRKVSDIGFTGKAVVDVDLNKCEPWLPAKASMGEKE
WYFFSQDRDKYPTGLRTNRATEAGYWKTTGKDKEIYRSGVLVGMKKTLLVFKGRAPKGEK
SNWVMHEYRLSKQPFNPTNKEEWVVCRVFEKSTAAKKAQEQPQSSQPSFGSPCDANSS
MANEFEDIDELPNLNSNSSTIDYNNHIHQYSQRNVYSEDNTTSTAGLNMNMNMMASTNLQS
WTTSLGPPPLSPINSLLLKAFQIRNSYSFPKEMIPSFNHSSLOQGVSNMIQNASSSSQVQ
PQPQEEAFNMDSIW*

>G1929 (1..366)

ATGTGTAGAGGCTTGAATAATGAAGAGAGCAGAAGAAGTGACGGAGGAGTTGCCGGAGT
CTCTGCACGAGACCGAGTGTTCGGTAAGGTGTGAGCTTTGCGACGGAGACGCCTCCGTG
TTCTGTGAAGCGGACTCGGCGTTCTCTGTAGAAAATGTGACCGGTGGGTTTCATGGAGCG
AATTTTCTAGCTTGGAGACACGTAAGGCGGTGCTATGCATTCTTGTGAGAACTCACG
CGCCGGTGCCTCGTCCGAGATCATGACTTCCACGTTGTTTTACCGTCCGTGACGACGGTC
GGAGAAACCACCGTGGAGAATAGAAGTGAACAAGATAATCATGAGGTTCCGTTTGTGTTTT
CTCTGA

>G1929 Amino Acid Sequence (domain in AA coordinates:31-53)

MCRGLNNEESRRSDGGGCRSLCTRPSVPVRCELCDGDASVFCEADSAFLCRKCDRWVHGA
NFLAWRHVRRVLCTSCQLTRRCLVGDHDFHVVLPSVTTVGETTVENRSEQDNHEVPFVF

L*

>G1930 (76..1077)

ATTCACATTACTAATCTCTCAAGATTTTCACAATTTTCTTGTGATTTTCTCTCAGTTTCTT
ATTTTCGTTTCATAACATGGATGCCATGAGTAGCGTAGACGAGAGCTCTACAACCTACAGAT
TCCATTCCGGCGAGAAAGTCATCGTCTCCGGCGAGTTTACTATATAGAATGGGAAGCGGA
ACAAGCGTGGTACTTGATTTCAGAGAACGGTGTCTGAAGTCGAAGTCGAAGCCGAATCAAGA
AAGCTTCCTTCTTCAAGATTCAAAGGTGTTGTTCCCTCAACCAATGGAAGATGGGGAGCT
CAGATTTACGAGAAACATCAACGCGTGTGGCTTGGTACTTTCAACGAGGAAGACGAAGCA
GCTCGTGTCTACGACGTCGCGGCTCACCGTTTCCGTGGCCGCGATGCCGTTACTAATTTT
AAAGACACGACGTTTCAAGAAGAGGTTGAGTTCTTAAACGCGCATTCGAAATCAGAGATC
GTAGATATGTTGAGAAAACACACTTACAAAGAAGAGTTAGACCAAGGAAACGTAACCGT
GACGGTAACGGAAAAGAGACGACGCGCTTGTCTTGGCTTCGATGGTGGTTATGACGGGG
TTTAAACGCGCGGAGTTACTGTTTGAAGAAAACGGTAACGCCAAGTGACGTCGGGAAACTA
AACCCTTTAGTTATACCAAAACACCAAGCGGAGAAACATTTTCCGTTACCGTTAGGTAAT
AATAACGTCCTCCGTTAAAGGTATGCTGTTGAATTTTCAAGACGTTAACGGGAAAGTGTGG
AGGTTCCGTTACTCTTATTGGAATAGTAGTCAAAGTTATGTGTTGACCAAGGTTGGAGT
AGATTCGTTAAAGAGAGAAGAGACTTTGTGCTGGTGATTTGATCAGTTTTTAAAGATCCAAC
GATCAAGATCAAAAATTCTTTATCGGGTGAAGATCGAAATCCGGGTTGGATCTAGAGACG
GGTCGGGTTATGAGATTGTTTGGGGTTGATATTTCTTTAAACGCGCTCGTTGTAGTGAAG
GAAACAACGAGGTGTTAATGTCGTCGTTAAGGTGTAAGAAGCAACGAGTTTTGTAATAA
CAATTTAACCACTTGGGAAAGAAAAAAGCTTTTTGATTTTAAATTTCTCTTCAACGTTA
ATCTTGCTGAGATTA

>G1930 Amino Acid Sequence (domain in AA coordinates: 59-124)

MDAMSSVDESSTTTDSIPARKSSSPASLLYRMGSGTSVVLDSENGVEVEVEAESRKLPS
RFKGVVPQPNGRWGAQIYEKHQVWLGTTFNEEDEAARAYDVAAHRFRGRDAVTNFKDTTF
EEVEEFLNAHSKSEIVDMLRKHTYKEELDQRKRNRDNGKETTAALASMVMTGFKTAE
LLFEKTVTPSDVGKLNRLVIPKHQAETHFPLPLGNNNVSVKMLLNFDVNGKVWRFRYS
YWNSSQSYVLTKGWSRFVKEKRLCAGDLISFKRSNDQDQKFFIGWKS KSGLDLETGRVMR
LFGVDISLNAVVVVKETTEVLMSSLRCKKQRL*

>G195 (51..1031)

TTTTCTTTTCTTCTTTTGGTTTAAAGTTTTTCTCTTTGTTCTTCGTCATGTCTCATG
AAATCAAAGATCTTAACAACTATCACTACACTTCATCGTATAATCATTACAATATCAACA
ACCAAAATATGATTAATCTCCCTTACGTTTCTGGTCCATCTGCTTATAATGCAAACATGA
TCTCATCATCACAAAGTAGGTTTTGATCTACCCCTCGAAGAACTTGAGTCCTCAAGGAGCCT
TCGAGTTGGGTTTTCGAGCTTTCTCCATCTCTTCTGACTTTTTTAAATCCTTCCCTCGATC
AAGAGAACGGTTTGTATAATGCTTATAATTATAATAGTAGTCAAAAGAGTCATGAAGTTG
TCGGTGATGGTTGTGCAACCATTAAAGAGTGAAGTTAGGGTTTCAGCATCTCCTTCTTCAA
GTGAGGCCGATCATCATCCAGGAGAAGATTCCGGCAAGATCCGGAAGAAAAGAGAAGTTC
CGGATGGAGGAGAAGATGATCAACGCTCTCAGAAAGTAGTTAAAACAAAGAAGAAAGAGG
AGAAGAAAAAAGAGCCACGAGTCTCGTTTCATGACTAAGACCGAAGTTGATCATCTCGAAG
ACGGCTATCGTTGGAGAAAGTATGGCCAAAAAGCAGTCAAAAACAGTCCTTATCCGAGGA
GTTACTATAGATGCACGACTCAGAAGTGCAACGTGAAGAAGAGAGTGGAGAGATCTTACC
AAGACCCAACGGTCGTCATCACAACTACGAGAGTCAACACAACCATCCGATCCCGACCA
ATCGTCGGACAGCAATGTTCTCTGGAACCCGATCTGATTATAACCCATCATCGTCTC
CAATATTCTCCGATCTCATCATCAATACTCCAAGAAGCTTCTCAAATGATGATCTCTTCC
GTGTGCCATACGCTAGTGTGAACGTGAACCTAGTTATCATCAACAGCAACATGGATTTC
ATCAACAGGAGAGTGAGTTCGAGCTCTTGAAGGAGATGTTTCTTCGGTTTTCTTCAAAC
AAGAGCCTTGATGATATAATATAATATAGAAACAATTTTTTTCTGCTAAGAAATATAGA
ACAAAACCTTGGATGCATAATAAGTGATGATAGTGTTATTTATTTTTCATGTATATATT
ATACATGTTTTTGTAACTAGCTATAGGATATACTGGTAGTAATTAAGCATAAATATGGAG
CCCTTCGACTTATTACAATAATTTTTGGTATGGAAGAAANTTNGNTACATGCCTGCCTTTT
NNNTTNNGG

>G195 Amino Acid Sequence (domain in AA coordinates: 183-239)

MSHEIKDLNNYHYTSSYNHYNINNNQNMNINLPVVGSPSAYNANMISSQVGFDLPSKNLSP
QGAFELGFELESPSSSDFFNPNSLDQENGLYNAYNINSSQKSHEVVGDGCATIKSEVRVSAS
PSSSEADHHPGEDSGKIRKKREVRDGGEDDQRSQKVVKTKKKEKKKEPRVSFMTKTEVD
HLEDGYRWRKYQKAVKNSPYPRSYRRTTQKCNVKKRVERSYQDPTVVITTYESQHNHP

IPNRRRTAMFSGTTASDYNPSSSPIFSDLIINTPRSFSNDDLFRVPYASVNVNPSYHQQQ
HGFHQESEFELLKEMFPSVFFKQEP*

>G1954 (196..1440)

ATTTATGACTTCTCAATACAAAAAGCTCCCTCACTTTTTTAAGTTTTGTCTTCTCTAAT
CCGTCTTCTTCTACTATCTTGCTGCTCTTTTATATACATCTCTCGTAAACCCCT
AGCAAATCATACAAGGTCAAGAAGCTTGACCTTCATTAGACTTAAGCAGTTTATAATCAA
CTACCACGAATAGCAATGGATAAAGATTACTCGGCACCAAACCTTCTTAGGTGAATCCTCA
GGCGGTAACGATGATAACAGCTCTGGTATGATAGACTATATGTTCAATAGAAACCTTCAA
CAACAACAAAAGCAATCGATGCCACAACAGCAGCAACATCAACTCTCTCCTTCCGGATTT
GGAGCAACACCCCTTGATAAAATGAACCTTCTCTGATGTGATGCAGTTTGGCGACTTCGGT
TCGAAACTTGCGTTGAACCAGACCAGAAACCAAGACGATCAAGAAACCGGGATTGACCCC
GTTTATTTCTTGAAGTTCCCTGTCTTGAACGACAAAATAGAGGACCATAACCAAACCCAA
CATCTCATGCCTTCTCATCAGACGTCTCAAGAAGGAGGTGAGTGTGGAGGAAACATAGGC
AATGTGTTTTCTTGAAGAAAAAGAAGATCAAGACGATGACAACGACAACAACCTCCGTGCAA
CTACGTTTTTATTGGAGGAGAAGAAGAAGATAGGGAGAACAAGAATGTTACGAAAAAGGAG
GTGAAGAGCAAGAGGAAGAGAGCTAGAACGAGCAAGACCAGCGAAGAAGTGGAAAGCCAA
CGGATGACTCATATCGCGGTGCAAGAAACCGTAGGAAGCAATGAATGAGCATCTTCGT
GTCCTTAGATCTCTCATGCCTGGCTCCTACGTTCAAAGGGGAGACCAAGCGTCAATCATA
GGAGGAGCAATAGAGTTTGTGAGAGAGCTCGAGCAACTCCTACAATGTCTTGAATCACAG
AAGCGTCGAAGAATCTTAGGAGAAACCGGTAGGGACATGACAACGACAACGACTTCTTCT
TCTTCTCCCATAACTACGGTAGCGAACCAGCACAAACCGCTCATTTATTACGGGAAATGTA
ACCGAGCTAGAGGGCGGAGGAGGGCTTCGGGAGGAGACTGCGGAGAACAAAGTCGTGCTTG
GCTGACGTGGAGGTGAAGCTGCTAGGGTTTGACGCCATGATCAAGATACTTTCAAGAAGA
AGGCCGGGACAGCTGATTAAGACTATAGCTGCTTTGGAGGATCTTCATCTCTCTATTCTT
CACACTAACATCACTACCATGGAACAAACCGTCTCTACTCCTTTAATGTCAAGATAACA
AGTGAACGAGGTTTACGGCAGAAGACATAGCAAGTCCATCCAACAGATATTTAGTTTC
ATTTCATGCAAAATACCAACATATCTGGAAGCTCTAACCTGGGAAATATTGTGTTTACTTGA
AAATCATCACACGGCGACAACCTTTGTACTGCTGGAAGATTACAGTACGTAATAATCTCT
ACATATTGGGTTTTATTCTCCAAGCATTTGGAAGAGTGTTTAAAGTTAAAGGGAGTGCTTA
CTTTATTTTTTTGGGGCTTTTTTTCATGCAATTTAAATTTTAGTGATGATTGTGTGCTTG
TAATGTTAGAACTCGTTGTTGTGATTCTGCTGCTTTGATTGTAGGTTTTGAACAAGCG
GTTTAGAATGCTAAACCACTTATTACTTGAAATAACTTTTTTCACAAAAAAGGAAAAA
AAGAAAAAA

>G1954 Amino Acid Sequence (domain in AA coordinates:187-259)
MDKDYSAPNFLGESSGNDNSSGMIDYMFNRNLQQQQKQSMPPQQQHQQLSPSGFGATPF
DKMNFSDVMQFADFGSKLALNQTRNQDDQETGIDPVYFLKFPVLNDKIEDHNQTOHLMPS
HQTSQEGGECGGNIGNVFLEEKEDQDDNDNNSVQLRFIGGEEEDRENKNVTKKEVKS
KRARTSKTSEEVESQRMTHIAVERNRRKQMNHLRLVRLSLMPGSYVQRGDQASIIIGGAIE
FVRELEQLQLCLESQKRRRILGETGRDMTTTTSSSSPITTVANQAQPLIITGNVTELEG
GGGLREETAENKSLADVEVKLLGFDAMIKILSRRRPGQLIKTIAALEDLHLSILHNTNIT
TMEQTVLYSFNVKITSETRTAEDIASSIQIFSFHANTNISGSSNLGNIVFT*

>G1958 (107..1336)

GTACCGTCGACCGATTATCCCCAAGAGGAGAATCCTCATAATCATTTTTCTCCGATTGAT
TCGCTCTTCTTGGTCTTGATTGCTTCATGAATTTCTAGGACAACAATGGAGGCTCGTCC
AGTTTCATAGATCAGGTTGAGAGACCTCACACGCACCTTCTCAATCCCATCTACAAAAA
ACCTTCACCCAGTAGAAGATAGTTTCATGAGATCAGATAACAACAGTCAGTTAATGTCTAG
ACCATTAGGACAAACCTACCATTTACTTTTCATCTAGTAACGGTGGAGCTGTTGGACATAT
ATGTTCTTCTTCATCATCTGGTTTTGCAACCAATCTCCATTACTCAACTATGGTATCTCA
TGAGAAAACAACAACACTACACAGGAAGCAGCAGTAATAATGCTGTGCAGACACCAAGCAA
CAACGATAGTGCTTGGTGTCTGATTCATTGCCAGGAGGGTTTCTTGACTTCCATGAAAC
CAACCCGGCGATTCAAAACAACCTGTGAGATTGAGGATGGTGGCATTGCGGCTGCTTTTGA
TGACATTCAAAACGAAGTGATTGGCATGAATGGGCTGACCATTGATCACTGATGATGA
TCCTTTGATGTCTACTAACTGGAATGATCTCTTGCTTGAAACAAATCCAATTCAGATTC
AAAGGACCAGAAGACACTGCAATTCGCAACCTCAGATTGTTTCAGCAGCAACCTTCTCC
GTCTGTGGAATTGCGACCTGTTAGCACAACATCTTCAAACAGCAATAACGGAACGGGCAA
GGCAGCAATGCGTTGGACGCCAGAGCTTCACGAGGCTTTGTTGAGGCTGTCAACAGTCT
TGGCGGTAGTGAAAGAGCTACTCTAAAGGGGTACTGAAGATTATGAAGTTGAAGGCTT

GACTATATATCATGTTAAAAGCCATTTACAGAAATATAGGACAGCTAGATATCGGCCAGA
ACCATCAGAAACTGGTTCGCCAGAAAGGAAGTTGACACCGCTTGAACATATAACATCTCT
TGATTTGAAAGGTGGGATAGGTATTACAGAGGCTCTACGACTTCAGATGGAAGTACAGAA
GCAACTCCATGAGCAGCTCGAGATTCAAAGAAACCTGCAACTCCGAATAGAAGAACAAGG
CAAGTACCTGCAAATGATGTTTCGAGAAGCAAACTCTGGTCTTACCAAAGGGACAGCCTC
AACATCAGATTCGGCAGCCAAATCTGAACAAGAAGACAAGAAGACTGCTGATTCGAAGGA
GGTTCAGAGAAGAAACCAGGAAATGTGAGGAACCTAGAATCTCCACAGCCAAAGCGTCC
CAAAATCGATAATTGAAAGTATTGGTCTTTTGTGCTGGATAATCTCGGAGTTTCAGAGTTAA
CAGTGATAGAGAGAACGAGCTCTTATCTTGAGGTTCTTCAGGACTTCTCTCGCGGCCGCT
CTAG

>G1958 Amino Acid Sequence (domain in AA coordinates: 230-278)

MEARPVHRSGSRDLTRTSSIPSTQKPSPVEDSFMRSDNNSQLMSRPLGQTYHLLSSSNGG
AVGHICSSSSSGFATNLHYSTMVSHEKQQHYTGSSSSNNAVQTPSNNSAWCHDSLPGGF
DFHETNPAIQNNCQIEDGGIAAADDIQKRSDWHEWADHLITDDPLMSTNWNLDLLETN
SNSDSKDQKTLQIPQPIVQQQSPSPVELRPVSTTSSNSNNGTGKARMRWTPELHEAFVE
AVNSLGGSERATPKGVLLKIMKVEGLTIYHVKSHLQKYRTARYRPEPSETGSPERKLTPL
HITSLDLKGGIGITEALRLQMEVQKQLHEQLEIQRLQLRIEEQKYLQMMFEKQNSGLT
KGTASTSDSAKSEQEDKKTADSKEVPEEETRKCEELESPQPKRPKIDN*

>G196 (111..1421)

TCGACATCAGATTTCTCTCACGGATTCTTAATCATTTTTATTATATTTGGATATTTGCTA
ATTCTTCCCGTGTATAAATCTCATATAAACACGCATCATACATATATATTATGTGCAGCG
TCTTTGAGTTTCAAGACATGGACAACCTCCAAGGAGATCTAACAGACGTCGTACGAGGAA
TAGGATCAGGCCACGTGTACCATCTCCTGGACCACCGGAAGGTCCATCTCCGAGCAGCA
TGTCTCCGCCGCCAACATCAGATCTCCACGTGGAATTCCTCCGCCGCTACTTCTGCCA
GCTGTCTCGCAAATCCCTTCGGAGACCCGTTTCGTAAGCATGAAGGATCCTCTCATCCACC
TCCCGGCCAGCTACATCTCCGGCGCCGGTGATAATAAAGCAACAAAAGTTTTGCAATCT
TTCCAAAGATTTTTGAGGATGATCATATTAAGAGTCAATGCAGTGTCTTCCCAAGAAATTA
AGATCTCGCAAAGTAACAATATCCACGATGCCTCCACGTGTAATCTCCGGCCATAACCG
TCTCCTCTGCCGCCGTAGCAGCTTCGCCGTGGGGCATGATCAACGTTAATACCCTAACAA
GTCCAAGAAACTGTTTACTTGTGCGATAATAATAACAACACGTCATCATGCTCACAGGTT
AGATCTCTTCTTCCCTCGGAATCTCGGAATTAAGAGAAGGAAGAGCCAGGCAAAGAAAG
TGGTGTGCATACCGCTCCAGCCGCTATGAACAGCCGGTCCAGTGAGAGAGTTGTTCCGT
CTGATCTATGGCTTGGCGAAAGTACGGTCAAAAACCTATCAAAGGTTCTCCTTATCCAA
GGGTTACTACAGATGTAGCAGCTCAAAGGTTGTTTACGCTAGGAAACAAGTCGAACGTA
GCCGCACTGATCCAAACATGTTAGTCATTACTTACACCTCTGAGCATAACCACCCATGGC
CTACTCAACGCAACGCTCTCGCAGGTTCCACTCGTTCCTCTTCTCCTCCTCTTTAAACC
CTTCTTCCAAATCCTCAACCGCAGCCGCCACTACTTCTCCCTCATCCAGAGTTTTCCAAA
ACAACAGCAGCAAAGACGAACCCAATAACTCCAACCTGCCTTCTCTTCCACTCATCCTC
CTTTTGACGCCGCCGCAATTAAGGAGGAGAACGTGGAAGAGCGTCAGGAAAAGATGGAGT
TCGATTATAATGACGTTGAAAATACCTATAGACCGGAGTTGTTGCAAGAGTTTCAACATC
AGCCGGAGGATTTCTTTGCCGATCTCGACGAGCTTGAGGGAGATTCTTTGACTATGTTGC
TCTCTCAGTAGCGCGGAGGCAACATGGAAAACAAAACGACGATTCCAGACGTTTTTA
GTGATTTCTTTGACGACGACGAGTCCTCAAGGTCGTTATAAATATTGTTGTTAATGTATA
CATAGAAATGAAATTATTCATGTAATTCGTTTTGTGTTAAATGACGGTATTTGCCTTTGC
A

>G196 Amino Acid Sequence (conserved domain in AA coordinates: 223-283)

MCSVFQMDNFQGDLDVVRGIGSGHVSPSPGPPEGSPSSMSPPPTSDLHVEFPSAA
TSASCLANPFGDPFVSMKDP LIHL PASYISGAGDNKSNKSF AIFPKIFEDDH IKSQCSVF
PRIKISQNNIHDA STCNPAITVSSAAVAASP WGMINVNTTNSPRNCLLV DNNNTSSC
SQVQISSSPRN LGIKRRKSQAKKVCIPAPAAMNSRSSGEVVP SDLWAWRKYGQKPIKGS
PYPRGYRCS SSGKSGSARKQVERSRTDPNMLVITYTSEHNHPWPTQ RNALAGSTRSSSSS
SLNPSSKSS TAAATSPSSRVFQNNSSKDEPNNSNL PSSSTHPPFDAAAIKEENVEERQE
KMEFDYNDVENTYRPELLQEFQHPEDFFADLDELEGDSL TMLLSHSSGGGNMENKTIP
DVFSDFDDDESSRSL*

>G1965 (1..609)

ATGGATAACTTCAATGTTGTTGCCAATGAAGACAATCAAGTGAATGATGTGAAGCCTCCA
CCACCCCCACCGCGAGTGTGTGCAAGATGTGATTCTGATAACACAAAATTTTGTACTAC

AACAATTATAGTGAGTTTCAACCGCGCTACTTCTGCAAGAACTGTCGAAGATACTGGACT
CATGGTGGGGCTTTAAGAAACGTACCAATTGGTGGGAGTAGTCGTGCCAAGCGGACAAGG
ATAAATCAACCTTCAGTTGCTCAGATGGTTTCTGTTGGAATCCAACCAGGGAACCGTTTT
AGTCTTTTGTCTCATATTCATGGTGGTATGGTAACAAATGTGCATCCAACCTCAAACCTTT
CGACCAAATCATCGCCTAGCTTTCCATAATGGATCATTGAGCAAGATTATTATGATGTT
GGGTCTGATAATCTTTTGGTAAACCAACAAGTTGGTGGATATGTTGATAATCACAACGGT
TATCAGATGAATCAAGTGGATCAATACAACTGGAACCAGAGCTTCAATAACGCTATGAAC
ATGAATTATAATAACGCTAGCACTAGCGGAAGGATGCATCCTAGTCATTTAGAGAAGGGT
GGTCCTTGA

>G1965 Amino Acid Sequence (domain in AA coordinates:27-55)
MDNFNVVANEDNQVNDVKPPPPPPRVCARCDSDNTKFCYNNYSEFQPRYFCKNCRRYWT
HGGALRNVPIGGSSRAKRTRINQPSVAQMVSVGIQGNRFSSLSHIHGGMVTNVHPTQTF
RPNHRLAFHNGSFQDYDVGSDNLLVNQQVGGYVDNHNGYHMNQVDQYNWNQSFNNAMN
MNYNNASTSGRMHPSHLEKGGP*

>G1976 (1..1152)

ATGACTGATCCTTATTCCAATTTCTTCACAGACTGGTTCAAGTCTAATCCTTTTCACCAT
TACCCTAATTCCTCCACTAACCCCTCTCCTCATCCTCTTCTCTGTTACTCTCCCTCT
TCCTTCTTCTTCTTCCCTCAATCCGGAGACCTCCGCGCTCCACCGCCGCCACCACTCCT
CCTCCTTCTCCTCCTCCTCGAGAAGCCCTCCTCTCCTCAGCCTCAGCCCCGCCAACAAA
CAACAAGACCACCATCACAAACCATGACCACCTTATTCAAGAACCACCTTCAACCTCCATG
GATGTCGACTACGATCATCACCATCAAGATGATCATATAACCTCGATGACGATGACCAT
GACGTCACCGTTGCTCTTCACATAGGCCTTCCAAGCCCTAGTGCTCAAGAGATGGCCTCT
TTGCTCATGATGTCTTCTTCTCCTCTTCTCCTCGAGGACCACTCATCATCAGAGGACATG
AATCACAAGAAAGACCTCGACCATGAGTACAGCCACGGAGCTGTGCGAGGAGGAGAAGAT
GACGATGAAGATTAGTCGCGCGGAGACGGCGGCTGTAGAATCAGCAGACTCAACAAGGGT
CAATATTGGATCCCTACACCTTCTCAGATTCTCATTGGCCCTACTCAGTTCTCATGTCTCT
GTTTGCTTCAAAACCTTCAACAGATACAATAACATGCAGATGCATATGTGGGGACATGGA
TCACAATACAGAAAAGGACCTGAATCTCTAAGGGGAACACAACCAACAGGAATGCTAAGG
CTTCCGTGCTATTGCTGCGCCCCAGGCTGTGCAACAACATTGACCATCCAAGGGCAAAG
CCTCTCAAAGACTTCAGAACCCCTTCAAACACATTACAAGAGAAAACATGGGATCAAACCT
TTCATGTGTAGGAAATGTGGAAGGCTTTCGCAGTCCGAGGGGACTGGAGAACACATGAG
AAGAATTGTGGCAAACCTTTGGTATTGCATATGTGGATCTGATTTCAAGCACAAGAGATCT
CTCAAAGATCACATCAAGGCTTTTGGGAATGGTCATGGAGCCTACGGAATTGATGGGTTT
GATGAAGAAGATGAGCCTGCCTCTGAGGTAGAACAATTAGACAATGATCATGAGTCAATG
CAGTCTAAATAG

>G1976 Amino Acid Sequence (domain in AA coordinates: 219-323)
MTDPYSNFFTDWFKSNPFHHPNSSTNPSPHPLPPVTPPSSFFFFPQSGDLRRPPPPPTP
PPSPPLREALPLLSLSPANKQQDHHHNDHLIQEPPSTSMVDVDYDHHHQQDHHNLDHDDH
DVTVALHIGLPSPSAQEMASLLMSSSSSRTHHHEDMNHKKDLDEHYSHGAVGGGED
DDEDSVGGDGGCRISRLNKGQYWIPTPSQILIGPTQFSCPVCFKTFNRYNNMQMHMWGHG
SQYRKGPESLRGTQPTGMLRLPCYCCAPGCRNNIDHPRAKPLKDFRTLQTHYKRKHGIXP
FMCRCGKAFVRGDWRTHKNCGLWYICGSDFKHRSKLDHIKAFGNHGHGAYGIDGF
DEEDEPASEVEQLDNDHESMQSK*

>G2057 (27..1289)

GCCGTCTCGACGAATATGCTCTACCAATGTCTGACGACCAATTCCATCACCCGCCGCTC
CTTCTTCAATGAGGCACCGTTCTACGTGCGATGCGCGGACGGCGGCTGCGGCGAGATTG
TTGAGGTGCAAGGTGGTCACATTGTTTGGTCTACCGGAAGAAAAGACCGCCACAGCAAAG
TCTGCACGGCTAAAGGGCCACGTGACCGGCGGTGAGACTCTCTGCTCACACGGCGATTCT
AGTTTTACGATGTTCAAGACAGGCTTGGTTTTCGACCGACCTAGCAAAGCCGTTGATTGGC
TTATCAAAAAGGCTAAGACTTCCATTGACGAGCTCGCTGAGCTTCTCCTCGGAATCCCG
CCGATGCAATTCGCCCTAGCCGCTGCTAACGCTAAACCCAGAAGAACCACCGCCAAAACCC
AAATCTCTCCGTCTCCGCCACCGCCGCAACAGCAACAACAACAACAGCTTCAGTTTCG
GTGTTGGCTTCAACGGAGGAGGAGCAGAGCATCCGAGTAACAACGAGTCGAGTTTTCTCC
CGCCGTCAATGGATTTCAGATTTCGATAGCTGACACTATAAAGTCGTTTTTCCGGTGATTG
GCTCTTCAACGGAGGCTCCTTTCGAATCATACCTTATGCACAACATATCATCATCAGCATC
CGCCGGATTGCTTCTCTGAACTAATAGCCAAAACCAAGATCTCCGTCTCTCGCTGCAAT
CGTTCCCGGATGGTCCACCGTCTGCTCTGCAACCACCAACATCACCACCACCTCTGCTT

CCGCCTCCGAGCCTACTCTGTTCTACGGACAGAGCAATCCGTTAGGGTTTGACACATCGA
 GTTGGGAGCAGCAGTCGTCGGAATTCGGAAGGATTAGAGACTAGTGGCTTGAACAGCG
 GCGGTGGCGGCGGAGCAACCGATACAGGAAACGGAGGAGGGTTTCTGTTCCGCTCCTCCTA
 CTCCTTCAACGACGTCGTTTCAGCCAGTTCTTGGCCAAAGCCAACAGCTTTATTCTCAGA
 GGGGTCCCCCTTCAGTCCAGTTACAGTCCCATGATCCGTGCTTGGTTTGATCCTCACCATC
 ATCACCATCCATCTCCACCGACGATCTCAACCACCACCATCACCTTCTCCACCGGTTTC
 ACCAATCAGCAATCCCCGGAATCGGATTCGCCTCAGGTGAATTCTCTTCGGGTTTTTCGCA
 TACCAGCAGGTTTTAGGGCCAAGAAGAGGAGCAGCAGCAGGTTCTACTCACAAGCCGT
 CCTCTGCTTCTCTATTTCTCGCCATTGACAATCGAACTAATCCTC

>G2057 Amino Acid Sequence (domain in AA coordinates: TBD)
 MSDDQFHHPPPPSSMRHRSTSDAADGGCGEIVEVQGGHIVRSTGRKDRHSKVCTAKGPRD
 RRVRLSAHTAIQFYDVQDRLGFDPRPSKAVDWLKKAKTSIDELAEPLPPWNPADAIRLAAA
 NAKPRRTAKTQISPSPPPPQQQQQQQLQFVGFGNGGAEHPSNNESSFLPPSMDSDSI
 ADTIKSFPPVIGSSTEAPSNHNLHMHYHHQHPPDLLSRTNSQNQDLRLSLQSFDPGPPSL
 LHHQHHTSASASEPTLFYQSNPLGFDTSSEWQSSSEFGRIQRLVAWNSGGGGGATDT
 GNGGGFLFAPPTPSTTSFQPVLGQSQQLYSQRGPLQSSYSPIRRAWFDPHHHHSISTDD
 LNHHLHPPPPVHQAIPGIGFASGEFSSGFRIARFQGEQEEQHDLGLTHKPSSASSISRH
 *

>G2107 (79..624)
 ACCACAAAACAGAGCAACACACAACAAAGCTTCATTTCAATTCTGTTTCGAGAACCCT
 TTGAGAACGAGATCGGAGATGGAACACGACGATACACCGTGGCGGAGATGAAGCCAAAG
 AAGCGTGTGGACGGAGGATTTTCAAGGAGACACGTACCCCAATCTACAGAGGCGTGCAGG
 CGTAGGGACGGCGACAAATGGGTATGCGAAGTCCGTGAACCGATTATCAGCGTCGAGTC
 TGGCTCGGAACCTATCCGACGGCAGATATGGCCGCACGTGCTCACGACGTGGCGGTTCTT
 GCTCTGCGCGGAGATCCGCGTGTGTTGAATTTCTCCGATTCTGCTTGGAGGTTGCCGGTG
 CCGGCATCCACTGATCCGGACACGATCAGGCGCACGGCGGCCGAAGCAGCGGAGATGTTTC
 AGGCCCGCGGAGTTTAGTACAGGAATTACGGTTTTACCTCAGCCAGTGAGTTTGACACG
 TCGGATGAAGGAGTCGCTGGAATGATGATGAGGCTCGCGGAGGAGCCGTTGATGTCGCCG
 CCAAGATCGTACATTGATATGAATACGAGTGTGTACGTGGACGAAGAAATGTGTTACGAA
 GATTTGTCACTTTGGAGTTACTAAAATACGTATGTGTTAAAAAACCAAGATCGTATGTG
 TATGTATGCATAATAAATGGGCTTAATGATGGGCATAGATATGATAGGTCCAGCCTATAT
 GTTAAATGTGTTTTATTTTTTGGTTTTATCTAGTTTCTAGGTATTTACCAAATTGTATTA
 GTATAAGTTTTATTAAAGAAATAATCAAAAATGTTGTTGCCAAAAAAAAAAAAAAAAAAAA
 AAAAA

>G2107 Amino Acid Sequence (domain in AA coordinates: TBD)
 MENDDITVAEMKPKKRAGRRIFKETRHPIYRGVRRRDGDKWVCEVREPIHQRRVWLGTYP,
 TADMAARAHDVAVLALRGRSACLNFSASAWRLPVPASTDPDTIRRTAAEAEMFRPPEFS
 TGITVLPSASEFDTSDEGVAGMMMLRLAEPLMSPPRSYIDMNTSVYVDEEMCYEDLSLWS
 Y*

>G211 (1..750)
 ATGATGTCATGTGGTGGGAAGAAGCCAGTGTCTAAGAAAACAACGCCGTGTTGCACGAAG
 ATGGGGATGAAGAGAGGACCATGGACGGTGGAGGAAGACGAGATTCTTGTGAGCTTCATT
 AAGAAAGAAGGTGAAGGACGGTGGCGATCGCTTCTAAGAGAGCTGGTTTACTCAGATGT
 GGAAAGAGCTGTCGTCTACGGTGGATGAATCTCCGACCCCTCGGTTAAACGTGGAGGA
 ATTACGTCGGACGAGGAAGATCTCATCTCCGTCTTCACCGCCTCCTCGGCAACAGGTGG
 TCATTGATCGCGGGAAGGATAACCGGAAGGACTGATAATGAAATTAAGAACTATTGGAAC
 ACTCATCTTCGTAAGAACTTTTAAGGCAAGGAATTGATCCTCAAAACCACAAGCCTCTT
 GATGCAAAACAACATCAATGATAAAGCAAGAAAGTTTCGGGTGGACAAAAGTACCCCTTA
 GAGCCTATTTCTAGTTCTCATACTGATGATACCACTGTTAATGGCGGGGATGGAGATAGC
 AAGAACAGTATCAATGCTTTGGTGGTGAACACGGCTACGAAGACTTTGGTTTCTGCTAC
 GACGACAAGTTCTCATGTTTCTTAATTCGCTCATCAACGATGTTGGTGATCCTTTTGGT
 AATATTATCCCAATATCTCAACCTTTGCAGATGGATGATTGTAAGGATGGGATTGTTGGA
 GCGTCGTCTTCTAGCTTAGGACATGACTAG

>G211 Amino Acid Sequence (conserved domain in AA coordinates: 24-137)
 MMSCGGKKPVSKKTPCCTKMGMRKPWTVEEDEILVSFIKKEGEGRWRSLLPKRAGLLRC
 GKSCRLRWNNYLRPSVKRGGITSDEEDLILRLHRLGNRWSLIAGRIPGRTDNEIKNYWN
 THLRKLLRQIDPQTHKPLDANNIHKPEEEVSGGQKYPLEPISSHTDDTTVNGGDGDS

KNSINVFGGEHGYEDFGFCYDDKFSSFLNSLINDVGDPFGNIIPISQPLQMDCKDGIVG
ASSSSSLGHD*

>G2133 (26..457)

ATCTCATCTTCATCCACCCAAAAACATGGATTCAAGAGACACCGGAGAACTGACCAGAG
CAAGTACAAAGGTATCCGTCGTCGGAAATGGGGAATGGGTATCAGAGATTCGTGTCCC
GGGAACCTCGTCAACGTCTCTGGTTAGGCTCTTTCTCCACCGCAGAAGGCGTGCCTAGC
CCACGACGTCGCTTTTTACTGCTTGACCCGACCATCTCCCTCGACGACGAATCTTTTAA
CTTCCCTCACTTACTTACAACCTCCCTCGCCTCCAATATATCTCCTAAGTCCATCCAAAA
AGCTGCTTCCGACGCCGGCATGGCCGTGGACGCCGATTCCATGGTGTGTGTCTGGGAG
TGGTGGTTGTGAAGAGAGATCTTCATGGCGAATATGGAGGAGGAGGACAACTTAGTAT
CTCCGTGTATGATTATCTTGAAGACGATCTCGTTTGATCTATACGAGTACGTTTTTAGCA
GTTAA

>G2133 Amino Acid Sequence (domain in AA coordinates:11-83)
MDSRDTGETDQSKYKGIIRRRKWKVSEIRVPGTRQRLWLGSFSTAEGAAVAHDVAFYCL
HRPSSLDDSFNFPHLTSLASNISPKSIQKAASDAGMAVDAGFHGAVSGSGGCEERS
MANMEEEDKLSISVYDYLEDDL*

>G2134 (36..644)

GAGCAAAAACCTTTGTGTGCGTGTGTGTGTGTGTTTCATGGCTGGTCTTAGGAATTCGGTA
ACAGCGACAAAGCGCAAAACGATGGCAAAGGTGTACCATCTGCCTACAGAGGAGTCCGGA
AGAGAAAATGGGGGAAATGGGTGTCTGAAATCCGTGAACCGGGGACCAAGAACCGTATCT
GGCTAGGCAGTTTCGAGACTCCTGAAATGGCTGCAACCGCATACGACGTGGCAGCATTTC
ATTTTCAGAGGGAGAGAAGCTCGTCTCAACTTCCCTGAGCTCGCCAGCAGCCTTCCACGTC
CTGCAGACTCTAGCTCAGACAGCATTCGCATGGCAGTTTCATGAGGCAACACTCTGCCGCA
CCACCGAAGGAACAGAGTCAGCCATGCAAGTGGACAGCTCAAGCTCCTCCAATGTAGCTC
CAACAATGGTCAGACTCTCGCCAGGAAATTCAGCGATCAACGAGTCAACTTTGGGAT
CTCCTACTACAATGATGCATTCAACATACGACCCTATGGAGTTTGCTAATGATGTGGAGA
TGAATGCTTGGGAAACATACAGAGTGACTTTCTTTGGGACCCCTAACCCCAAAACCTAA
CTCATGGAGAGCTTCTACAGCTCAATCTTACAATACCAGCATAAGTTACTGGCTTAGAAT
ACTTAAATTTATTGAAGTTTAGTTTTCAGAGTCTACCACAAGGGTTGTTGATTCTGACGT
TATAGCAAAGAATAAAGCTCATCAGATTTTGGAGGAAAGACTCTATGAGCTTGATGGGT
CCCTGAAAGGACCTCTTCACAAATATTTTAAATTTTGTGTTACTAGTAGAAACATAGA
TTATGAGGTGTGACTTATTATTATTTTACAATTGTTGTACCTCATTGATGTATTTG
ATTT

>G2134 Amino Acid Sequence (domain in AA coordinates: TBD)
MAGLRNSGNSDKAQNKGKVPVPSAYRGVRRKWKVSEIREPGTKNRIWLGSFETPEMAA
TAYDVAAFHFRGREARLNFPELASSLPRPADSSSDSIRMAVHEATLCRTTEGTESAMQVD
SSSSSNVAPTMVRLSPREIQAINESTLGSPTTMMHSTYDPMEFANDVEMNAWETYQSDFL
WDP*PQNLTHGELLQLNLTPA*

>G2151 (236..1321)

TTTTTTTTTTAGGGTTCATAAGAACAATTGGATTTTGAGCTCACAGTATAAATAACCCG
ACTTTGATTACTGGTAATTTTAAACCGCCATTGTTGTTCTCTTACTACTTTTGGGAA
TTAGGGTTTATGATTTCTGGGTATTAGATTAGATAAATTTGTTTCTTTTGTAAATC
AATTTAAAAATCTCTTATTTCTGTTAAAGACTTGTAATTTTGGAGTTTAAATGCATGGA
CGGAAGAGAAGCAATGGCATTTCAGGCTCGCATTCTCAGTACTATCTTCAAAGAGGAGC
CTTTACTAATCTCGCACCTTCCCAAGTCGCGAGTGGGCTTCACGCGCCGCCACATAC
GGGATTGAGGCCAATGTCTAACCCCTAACATTATCACCCTCAGGCTAACATCCAGGACC
TCCTTTCTCGGATTTTGGACACACCATTCACATGGGAGTGGTCTCCTCTGCTTCTGATGC
TGATGTGCAACCGCCACCGCCACCGCCACCAGAGGAACCGATGGTTAAGAGGAAACG
TGGACGGCCAAGAAAGTATGGAGAACCGATGGTTAGTAATAAGTCTAGGGACTCTTCTCC
AATGTCGTATCCTAATGAACCTAAACGGGCCAGAGGTCGACCTCCTGGAACCTGGAAGGAA
GCAACGCTTGGCTAATCTTGGTGAGTGGATGAATACTTCAGCTGGACTTGCTTTTGACAC
TCATGTGATCAGCATTTGGAGCAGGAGAAGACATTGCTGCGAAAGTTTGTGCTTTTACCA
ACAAAGACCTCGGGCTCTTTGTATAATGTGAGGCACTGGAACATTTCTTCAGTCACTCT
GTGCAAAACCGGTTCAACCGATCGTCACTTAACATACGAGGGACCTTTTGAGATTATAAG
TTTTGGTGGATCTTATTTGGTGAATGAAGAAGGTGGATCCAGAAGTCGAACAGGCGGATT
GAGTGTCTCTCTTCTCGTCCCGATGGTAGTATTATGCGGTTGGAGTTGACATGCTTAT
CGCAGCCAACCTTGTTCAGGTGGTGGCATGTAGTTTTGTATACGAGCAAGGGCAAAGAC

TCATAATAACAATAACAAGACCATCAGACAAGAAAAGGAACCAATGAAGAGGACAACAA
TAGTGAAATGGAGACCACACCGGGTAGTGCAGCTGAACCAGCAGCATCTGCGGGTCAGCA
GACGCCACAGAACTTCTCTTCTCAGGGAATAAGGGGGTGGCCCGGTTTCAGGCTCAGGCTC
TGGCAGATCACTTGACATTTGCAGAAACCCACTCACTGATTTTGATTGACTCGTGGATG
ATATACACTATTAGTCTTTGAAGCAGCAGCATACAAAATGTGATTGCTGTACATATGTTA
TTGTAGATTCTCTCTGGAATGTTGAAATCAGACATTTAAGGATTGATACTAGATCTCT
CAGCTCCTTCTAACATTGTTAATGTAACAGAACCCTCCCACTTTTCATGCTATTTGC

>G2151 Amino Acid Sequence (domain in AA coordinates:93-113, 124-144)

MDGREAMAFPGSHSQYYLQRGAFNLAPSQVAGSLHAPPPHTGLRPMSPNIHHPQANNP
GPPFSDFGHTIHMVSSASDADVQPPPPPPPEEPMVKRKRGRPRKYGEPMVSNKSRDS
SPMSDPNEPKRARGRPPTGRKQRLANLGEWMNTSAGLAFAPHVISIGAGEDIAAKVLSF
SQQRPRALCIMSSTGTISSVTLCKPGSTDRHLTYEGPFEIISFGGSYLVNEEGGSRRTG
GLSVLSRPDGSIIAGGVDMLIAANLVQVVACSFVYGARAKTHNNNNKTIHQEKEPNEED
NNSEMETTPGSAEPAASAGQOTPQNFSSQGIKRWPGSGSGSGSLDICRNPLTDFDLTR
G*

>G2154 (82..1317)

GCAAAAAGAAAAAATGAAAAAATCCCTAACTCTCTCTCTCTAGAAATTCTTATTTTGG
TGCGTATCTCTCTAAAAAGGAATGGATCCTAACGAAAGCCACCATCACCACCAACAACAA
CAGCTCCATCACCTCCACCAACAGCAACAGCAACAGCAGCAGCAACGACTCACTTCT
CCTTACTTCCACCACCAACTACAGCACCATCACCACCTTCCAACCACCGTAGCAACCACC
GCTTCTACCGGAAACGCCGTTCCATCTTCCAACAATGGGCTTTTCCCTCCGCAGCCTCAG
CCACAGCACCAGCCTAATGATGGGTCACTTCTCTCGCGGTGTACCCTCATTCACTTCCG
TCCTCGGCTGTGACGGCGCGGATGGAGCCGGTAAAGAGGAAGAGGGGTCGACCAAGAAAG
TATGTGACGCCGGAACAAGCCCTAGCGGCTAAGAAATTGGCGTCTTCTGCGAGTAGTTCG
TCTGTCTAAACAGAGCGAGAGCTTGTCTGTGTACCGGTGGTACGGTATCGACTAATTCC
GGGTCACTCAAGAAATCTCAGCTTGGTCTGTCTCGGGAACCTGGACAATGTTTACTCCG
CATATTGTTAATATAGCTCCTGGCGAGGATGTGGTCCAGAAAATTATGATGTTGCAAAAC
CAAAGCAAGCATGAATATGCGTCTTCTTCTGCATCAGGCACTATCTCTAATGCATCCTTG
CGCCAACCGGCTCCATCAGGAGGCAACTTACCATATGAGGGTCAATACGAGATTCTCTCA
CTATCTGGATCTCTATATCCGAACCTGAACAAGGTGGTAAATCCGGCGGCCTTAGCGTTTCT
TTATCTGCTTCAGATGGTCAGATCATCGGTGGAGCGATTGGTAGCCATCTCACAGCTGCT
GGCCCCGTTTCAGGTGATTCTTGGTACGTTTTCAGCTTGATAGAAAGAAGGATGCCGCCGGG
AGTGGTGGGAAAGGGGATGCTTCAAACAGTGGAAGTCGGTTAACTTCTCCTGTAAGCTCT
GGACAGTTGCTTGGCATGGGTTTCCCTCCTGGTATGGAATCTACGGGAAGAAATCCAATG
AGGGGAAACGACGAGCAACATGATCATCATCATCAAGCCGGTTTGGGTGGACCTCAT
CATTTTCATGATGCAAGCGCGCGAGGGGATACACATGACACATTCCAGGCCATCTGAATGG
CGCGGAGGAGCAACAGCGGTATGATGGCAGAGGCGGTGGCGGGTATGATTTGTGAGGA
AGGATAGGACATGATCGTTCGGAGAATGGAGATTACGAGCAGCAATACCGGATTAGCAG
AGCTTCCAGGAGAAGTGTGTAGAGTTTAGATCCCAAGTAGAGAAACAGAAGCGAGCAAA
GAATCTGAAGTGAAGAGGACTTATTAGACAGAGACTCGTCTGAAGGGTCTTTAATCATA
GAAAGAAGTTGCTGAGTGATTGCTTTTGTCTTCTTCTTGGTACGGTGTATTATATTAAC
TCCACAACCTTTTTTTTATACTTTTCAGTAACGATTCTCCTTCACTTTCAATTTTATTCTT
TTTTTTTATACTCTTTTTCTTTCTTATAATATTTTTTTTGGTTTTTCTTTCGTTTGTTA
CTAAAAAGGAAATGCTCTTTTGTGAAATATATACACTTCGTTTG

>G2154 Amino Acid Sequence (domain in AA coordinates:97-119)

MDPNESHHLHQQQLHLHQQQQQQQQQLTSPYFHHQLQHHLHLPPTVATTASTGNAV
PSSNNGLFPPQPQPHQPNDSGLAVYPHVPSSAVTAPMEPVKRRGRPRKYVTPEQA
LAAKKLASSASSSSAKQRRELAAVTGGTVSTNSGSSKKSQLGSVGKTGQCFTPHIVNIAP
GEDVVQKIMMFANQSKHELCLVLSASGTISNASLRQPAPSGGNLPYEGQYEILSLSGSYIR
TEQGGKSGGLSVLSASDGQIIGGAIGSHLTAAGPVQVILGTFQLDRKKDAAGSGGKGD
SNSGSRLTSPVSSGQLLGMFPFGMESTGRNPMRGNDQHDHHLHQAAGLGGPHHFMQAP
QGIHMTSRPSEWRGGGNSGHDGRGGGGYDLSGRIGHSESENGDYEQQIPD*

>G2157 (306..1238)

TCCTTTGATTTTAACTTTTTCAGTAGCAAGCCAAAAAAGGACAGACAAAGAAGTT
CCTTTTATGATAAAGGTATGATGATAGCAAAACAAATGATACCCCATGTCTTGTGTGTCT
GCTTCATGCAACATGTTGGTTTGGATTGGTTAATCTAAAAGTTTAAGATAAGGTTTTCG
GATTCTCTTCTGTCTTGTAAATAGTTTCTTGTCTCGAGAGCCATCAACACCAACTTCAACA

AAAAAACAAGAAAAAGAAAAAGATTCTCTTTCTCGTTTTATTTCCATTAGAGAAGAAAA
 AAAGAATGGCGAATCCTTGGTGGGTAGGGAATGTTGCGATCGGTGGAGTTGAGAGTCCAG
 TGACGTCATCAGCTCCTTCTTTGCACCACAGAAACAGTAACAACAACCAACCCACCGACTA
 TGACTCGTTCCGATCCAAGATTGGACCATGACTTCACCACCAACAACAGTGGAAGCCCTA
 ATACCCAGACTCAGAGCCAAGAAGAACAAGACAGCAGAGACGAGCAACCAGCTGTTGAAC
 CCGGATCCGGATCCGGGTCTACGGGTCGTCGTCCTAGAGGTAGACCTCCTGGTTCCAAGA
 ACAAAACCAAGAGTCCAGTTGTTGTTACCAAGAAAGCCCTAACTCTCTCCAGAGCCATG
 TTCTTGAGATTGCTACGGGAGCTGACGTGGCGGAAAGCTTAAACGCCCTTTGCTCGTAGAC
 GCGGCCGGGGCGTTTCGGTGCTGAGCGGTAGTGGTTTGGTTACTAATGTTACTCTGCGTC
 AGCCTGCTGCATCCGGTGGAGTTGTTAGTTTACGTGGTCAGTTTGAGATCTTGTCTATGT
 GTGGGGCTTTTCTTCTTACGTCTGGCTCTCCTGCTGCAGCCGCTGGTTTAACCATTTACT
 TAGCTGGAGCTCAAGGTCAAGTTGTGGGAGGTGGAGTTGCTGGCCCGCTTATTGCCTCTG
 GACCCGTTATTGTGATAGCTGCTACGTTTTCGAATGCCACTTATGAGAGGTTACCGATTG
 AGGAAGAACAACAGCAAGAGCAGCCGCTTCACTAGAAGATGGGAAGAAGCAGAAAGAAG
 AGAATGATGATAACGAGAGTGGGAATAACGAAACGAAGATCGATGCAGCCGCCGATGT
 ATAATATGCCTCCTTAATTTTATCCCAAATGGTCATCAAATGGCTCAACACGACGTGTATT
 GGGGTGGTCTCCTCCGCTCGTGCTCCTCCTTCGTATTGATTAGTTAGATAGGCGGTGGTTG
 GTGCGTTCTTTTACTGGAATGATTATATTTTCCATTAGGATGGTTAGGCTTTTGTATTAT
 TAAAGCTATCAAGTTTCTTTTTCGATAATTCGGATGACAATTAGCTAGTGTTT
 GTTGTGTTGTTTGTGGCGCTTTTCTGACTTGACTATTTTGTATCGCGGATAGCTTTGTA
 TGAAAGTGAATTGATTGTAGAATCGTCTTTTGAATTTTGATGTTGAAAAAACCAA
 >G2157 Amino Acid Sequence (domain in AA coordinates: 82-102, 164-107)
 MANPWVGNVAIGGVESPTSSAPSLHHRNSNNNPPTMTRSDPRLDHDFTTNNSGSPNT
 QTQSQEEQNSRDEQPAVEPGSGSGSTGRRPRGRPPGSKNPKSPVVVTKESPNLSQSHVL
 EITAGDVAESLNAFARRRGRGVSVLSGSLVTNVTLRQPAASGVVSLRGQFEILSMCG
 AFLPTSGSPAAAAGLTIYLAGAQGVVGGVAGPLIASGPVIVIAATFCNATYERLPIEE
 EQQQEQPLQLEDGKKQKEENDDNESGNNNGNEGSMQPPMYNMPNFI PNHGQMAQHDVYWG
 GPPPRAPPSY*
 >G2181 (1..1005)
 ATGATGCTTGCGGTGGAAGATGTGTTAAGCGAACTCGCCGGAGAAGAAAGGAACGAGAGA
 GGATTGCCACCTGGCTTCCGGTTTCACCCGACGGACGAAGAGCTCATTACCTTCTACTTA
 GCTTCCAAATCTTCCATGGTGGTCTCTCCGGCATTACATTTCGAAGTTGATCTCAAC
 CCTGTGAACCTTGGGAGCTTACCAGAAATGGCGAAGATGGGAGAGAGAGTGGTACTTT
 TATAGTCTAAGGGACAGGAAATATCCGACAGGTTTGAGGACTAACAGAGCAACTACTGCT
 GGATACTGGAAGCTACCGGCAAAGATAAGGAAGTCTTCTCCGGCGGAGGAGGACAGCTT
 GTTGGGATGAAGAAGACGTTGGTGTCTACAAAGGTAGGGCTCCACGTGGCCTCAAGACT
 AAGTGGGTGATGCATGATGATATCGCCTCGAAAACGACCATTACACCGCCACACGTGTAAG
 GAGGAATGGGTGATTTCAGAGTGTTCAATAAAACAGGAGACAGAAAAAATGTTGGATTA
 ATCCATAACCAAATCAGCTACCTTCATAACCATTCCTCTCAACAACACATCATCATCAT
 CATGAAGCCTTACCTTTGCTTATAGAACCCTTCCAACAAAACCCTAACCAACTTCCCATCA
 CTACTCTACGATGATCCACACCAAACTACAATAATAACAACCTCCTTCATGGATCATCA
 GGCCACAACATCGACGAGCTCAAAGCCTTAATCAACCCTGTGCTCTCTCAGCTCAACGGT
 ATCATCTTCTCCTTCAGGGAACAACAACGACGAAGACGACTTCGACTTTAACCTCGGC
 GTGAAAACAGAGCAGTCTTCAACGGTAACGAAATTGACGTACGAGATTACTTGGAGAAC
 CCTCTGTTTCAGGAAGCGAGTTATGGTCTGTTGGGTTTTTCTGCTCTCTCCTGGACCTCTT
 CACATGCTACTAGATTCTCCATGTCTTTAGGATTCCAGCTGTAG
 >G2181 Amino Acid Sequence (conserved domain in AA coordinates:22-169)
 MMLAVEDVLSELAGEERNERGLPPGFRFHTDEELITFYLASKIFHGGLSGIHISEVDLN
 RCEPWELPEMAKMGEREWYFYSRLDRKYPTGLRTNRRATTAGYWKATGKDKEVFSGGGQQL
 VGMKKTLLVFYKGRAPRGLKTKWVMHEYRLNDRHSHRHTCKEWWICRVFNKTGDRKNVGL
 IHNQISYLNHNSLSTTHHHHHEALPLLEPSNKTLTNFPSSLYDDPHQNYNNNNFLHGSS
 GHNIDELKALINPVVSQNLNGIIFPSGNNNNDEDDDFNGLGVKTEQSSNGNEIDVRDYLEN
 PLFQEASYGLLGFS SSPGLHMLLDSPCLPLGQL*
 >G221 (115..795)
 CTCTCTTATTCTCTCACTCTTTTTTTTTTATATTCTCTCTCTCTAAATCTATAAAATAT
 ATTTAAAAAATTGATCGTATATAATAAAGTAAATAAAGAATAATAACAAAAAAATGGAG
 AAAAGAGGAGGAGGAAGTAGTGGAGGTTCCGGATCATCAGCAGAAGCAGAAGTGAGAAAA

GGACCATGGACGATGGAAGAAGATCTTATTCTTATCAACTATATCGCCAACCACGGCGAT
GGTGTGTTGGAATTCTCTCGCCAAATCTGCAGGTCTAAAACGAACCGGGAAAAGTTGCCGG
CTCCGGTGGCTGAAC'TATCTCCGCCCCGACGTACGACGGGGAAACATCACTCCAGAAGAG
CAACTTATCATCATGGAAC'TTCATGCTAAGTGGGGAAACAGGTGGTTCGAAAATCGCCAAA
CATCTTCCAGGAAGAACGGACAACGAGATCAAAAATTTCTGTAGGACAAGAATTCAAAAA
TACATCAAGCAATCGGATGTAACAACAACATCGTCCGTTGGATCTCATCATAGCTCAGAG
ATCAACGATCAAGCTGCAAGCACGTGAGCCATAATGTCTTTTGTACACAAGATCAAGCG
ATGGAGACTTAT'TCTCTACACCGACATCATATCAACATACCAATATGGAATTCAACTAT
GGTAAC'TATTCGGCCCGGGCAGTGACGGCAACCGTGGATTATCCAGTACCGATGACCGTT
GATGATCAAACCGGTGAAAAC'TATTGGGGCATGGATGATAT'TTGGTCATCAATGCATTTA
TTGAATGGTAAT'TGATGATCGGTGGACAAAACATGGAATATTAATTGAGTATTATATAT
GATTTT'TAGGAGTACTATTATTAGTACGTGACATGTATATGTTTTGCTCGTTGTAGAG
GTTTGGGGTTATAATTAATATATAATGTTATCTAATATGCAACCTTGATACATATTGGA
TCTTTATTGAACCATGTTATACATAAAATAAAATGTTGAAGGGTGCATAAAAAAAAAA
AAAAAAAAAAAAA

>G221 Amino Acid Sequence (domain in AA coordinates: 21-125)
MEKRGSSSSGSSGSSAEVRKGPWTMEEDLILINIANHGDGVWNSLAKSAGLKRTGKS
CRLRWLNLRPDVRRGNITPEEQLIIMELHAKWGNRWSKIAKHLPGRTDNEIKNFCRTRI
QKYIKQSDVTTTSSVGSHTSSEINDQAASSTSHNVFCTQDQAMETYSPTPTSYQHTNMEF
NYGNYSAAAVTATVDYPVPMVTDDQTGENYWGMDIWSMHLNNGN*

>G2290 (119..982)

TTCTTTCTTTCTTTCTTTCTTTCTTTCAATCAAGAACAACCCCTAGCTCCTCTCTTTTCTC
TCTCTACCTCTCTTTCTCTATCTTCTCTTATCACTACTTCTCTCGCGGATCAATCATCAT
GAACGATCCTGATAATCCCGATCTGAGCAACGACGACTCTGCTTGGAGAGAACTCACACT
CACAGCTCAAGATTCTGACTTCTTTCGACCGAGACACTTCCAATATCCTCTCTGACTTCGG
TTGGAACCTCCACCACTCCTCCGATCATCTCACAGTCTCAGATTGACTCCGATTTAAC
ACAAACCACCGGAGTCAAACCTACCACCGTCACTTCTTCTTGTTCCTCATCCGCCGCCGT
TTCCGTTGCCGTTTACCTCTACTAATAATCCCTCAGCTACCTCAAGTTCAAGTGAAGA
TCCGCCCGAGAAGTCAACCGCCTCCGCCGAGAAAACACCACCACCGGAGACACCAAGAG
GGAGAAGAAGAAGGCTCAAAGCGAATTCCGGCAACCAAGATTGCGATTTCATGACCAAGAG
TGATGTGGATAATCTTGAAGATGGATATCGATGGCGTAAATATGGACAAAAAGCCGTCAA
GAATAGCCCATTCCTCAAGGAGTACTATAGATGCACAAAACAGCAGATGCACGGTGAAGAA
GAGAGTAGAACGTTTCATCAGATGATCCATCGATAGTGATCACAACATACGAAGGACAACA
TTGCCATCAAACATTGGATTCCCTCGTGGTGGAACTCTCACTGCACACGACCCACATAG
CTTCACTTCTCATCATCATCTCCCTCCTCCATTACCAAATCCTTATTATTACCAAGAAGT
CCTTTCATCAACTTTCACAGAGACAATAATGCTCCTTACCGCGGTTACCCCGACCTACTAC
TGAAGATACACCTGCCGTGTCTACTCTCATCAGAGGAAGGCTTACTTGGTGATATTGTACC
TCAAATATGCGCAACCCCTTGAGGTAAGCTTGGTACGTAGCAATAGCTAAGGAGGTGCTA
ACTCATTATATATAGAAGATATTGCAGACCAGAATATGCGCAGGGAGGGTATAACAATAT
GGCGTTGTAACAATGGATCTATATATTACCTCATTGTTGATCAATAGCACACCACCGGTA
CGTTTGCAATTTCTTCATGTATATTTCTTGTATATATGTAGTTATATATCCAGGTATAA
TTTTGATGTAACACAACATTAATCTTAATCGTGATCCATCCACATTTGATGCATGTAT
GTGCACTTAAGAAAAAGAACATGGAGGAAATAACGTTATTTTTTATTATTCT

>G2290 Amino Acid Sequence (conserved domain in AA coordinates:147-205)

MNDPDNPDLSNDDSAWRELTLTAQDSDFDRDTSNLSDFGWNLHSSDHPHSLRFDSDL
TQTTGVKPTTTSVSSSSAAVSVAVTSTNNNPSTSSSSSEDPAENSTASAEKTPPPETPV
KEKKKAQKRIRQPRFAFMKSDVDNLEDGYRWRKYGQKAVKNSPFRSYRCTNSRCTVK
KRVERSSDDPSIVITTYEGQHCHQTIGFPRGGILTADPHSFTSHHHLPPPLPNPYYYQE
LLHQLHRDNNAPSRLRPRTTEDTPAVSTPSEGLLDIVPQTMARNP*

>G2299 (231..941)

GCCAAAATTTTACCAACATTTTCTCTCTCATATCAAAGTTTCTCTCTCATTTCTTCAT
CACACTTCACTGCCCTGTTTTTTTTCTCATTTTGAATAGTTCTCAAACCTTATATATTTT
TCCCCCTGAAGCCTAGCTATTTCTTTTATTTGCATTAATCTCGGGATCCGAATCGAAAA
AAGCAATCAGAATAATAGACTTGTACGATACTTGTGCCAAGCTAACACAATGGCAGAGG
AATACTACAGCCTCCGCTCGGAGAGAGTAAGTCAAGCTTCTGTCCCTAACTCGGAGTCTG
ACTCAGTGAGTGACAAAAGCAAAGCTGAGCAAAGCGAGAAGAACTAAACGTGGGAGAG
ACTCCGGTAAACACCCTGTTTATCGCGGAGTAAGGATGAGGAAGTGGGGAAAATGGGTGT

CGGAGATTCTGTGAGCCGAGGAAGAAATCACGTATTTGGCTGGGAACTTTCCCGACGCCGG
AGATGGCGCGCGCTGCACACGACGTGGCGGCTCTGAGCATTAAGGAACGGCCGCTATAC
TAAACTTCCCTGAACCTCGCTGACTCATTCCCTCGACCCGTTTCATTAAGCCCTCGAGACA
TTCAGACAGCAGCTCTTAAAGCAGCTCACATGGAACCGACGACGTCGTTTTCATCTTCCA
CGTCTTCGTCTCGTCTTTGTCTTCTACGTCTTCGCTCGAGTCTCTTGTGTTGGTGATGG
ACCTCTCGAGGACTGAGTCGGAGGAGCTCGGTGAGATTGTGGAGCTTCCAAGTCTCGGGG
CGAGTTACGACGTCGACTCGGCTAACCTTGGGAACGAGTTTGTCTTCTATGACTCAGTTG
ACTACTGTTTATATCCGCCCGCTGGGGACAGTCGTCGGAAGATAACTATGGTCACGGAA
TTAGCCCTAATTTTGGCCATGGCTTGTCTATGGGATCTCTAACAGTTTATTTTGTATCATT
ACCATAATGTTTTGTATAAACAGTTTATTTTGTATCATTGCCATAATGTTTTGTTAAT
CACGTTTTTAAACCCCTTGTCTGTTTTTGTGTTTTTTTTTGTAGTTTTT
>G2299 Amino Acid Sequence (conserved domain in AA coordinates:48-115)
MAEEYYSLSRERVQLVLPNSES DSVSDKSKAEQSEKTKRGRDSGKHPVYRGVVRMRNWG
KWFSEIREPRKKSRIWLGTFPTPEMAARAHDAALSIKGTAAILNFPPELADSFPRPVLS
PRDIQTAALKAAHMEPTTSFSSSTSSSSSLSSSTSSLESILVLMDLSTRTESEELGEIVELP
SLGASYDVDSANLGNFVFYDSVDYCLYPPPWGQSSSEDNYGHGISPNFGHLSWDL*
>G2340 (274..1275)
ATACAAAACCTCCCTCTTCTCTATCTTCTTCATCTTAAAGAAAAAATAAGAGATATTCGTA
AAGAGAGAACACAAAATTTTCAGTTTACGAAAAGCTAGCAAAGTCGAGTATCGAGGAATAA
CAGAATAAGACGTATCTATCCTTGCCCTAATGTTCTTACCAAAGATCTAGTCCTTTCTT
TGTATGATCGATCCATCACAAAGCCCAACAACAACAACTACATCTCTTCTCTATCTCT
AGCTTCTATTTTTTAATACATTCAAGAATCAAGAATGGTACGGACGCCGTGTTGTAGAGCA
GAAGGGTTGAAGAAAGGAGCATGGACTCAAGAAGAAGACCAAAGCTTATCGCCTATGTT
CAACGACATGGTGAAGGCGGTTGGCGAACCTTCCGGACAAAGCTGGACTCAAAGATGT
GGCAAAGCTGCAGATTGAGATGGGCGAATTACTTAAGACCTGACATTAAACGTGGAGAG
TTTAGCCAAGACGAGGAAGATTCCATCATCAACCTCCACGCCATTTCATGGCAACAAATGG
TCGGCCATAGCTCGTAAATACCAAGAAGAACAGACAATGAGATCAAGAACCATTGGAAC
ACTCACATCAAGAAATGTCTGGTCAAGAAAGGTATTGATCCGTTGACCCACAAATCCCTT
CTCGATGGAGCCGGTAAATCATCTGACCATTCCGCGCATCCCGAGAAAAGCAGCGTTTCAT
GACGACAAAGATGATCAGAAATTCAGAAATCAAAAGTTGTGAGGATCATCATCAGCTCGG
TTTTTGAACAGAGTAGCAAACAGATTCCGGTCATAGAATCAACCACAAATGTTCTGTCTGAT
ATTATTGGAAGTAATGGCCTACTTACTAGTCACACTACTCCAAC TACAAGTGTTCAGAA
GGTGAGAGGTCAACGAGTTCCTCTCCACACATACCTCTTCAATCTCCCATCAACCGT
AGCATAACCGTTGATGCAACATCTCTATCCTCATCCAGTTCCTGACTCCCCGACCCG
TGTTTATACGAGGAAATAGTCGGTGACATTGAAGATATGACGAGATTTTCATCAAGATGT
TTGAGTCATGTTTATCTCATGAAGATTTATTGATGTCCGTTGAGTCTTGTGTTGGAGAAT
ACTTCATTCATGAGGGAATTACAATGATCTTTCAAGAGGATAAAATCGAGACGACGTCG
TTTAATGATAGCTACGTGACGCGCATCAATGAAGTTGATGACTCCTGTGAAGGGATTGAC
AATTATTTTGGATGAGTTATATTGATGATGATGAAAATTTGCATTGGCATGTAAATCAA
TTAGAGTTTGATTTGCTATGGTGTGTTTTAGTTTGTGTGTGTAGTGTGTTTCGACCGTCAA
AAAAAAAAAAAAAAAAAAAAAAAAA
>G2340 Amino Acid Sequence (domain in AA coordinates:14-120)
MVRTPCRAEGLKKGAWTQEDQKLIAYVQRHGGGWRTLPDKAGLKRCGKSCRLRWANY
LRPDIKRGEFSQDEEDSIINLHAIHGNKWSAIARKIPRRTDNEIKNHNWTHIKKCLVKKG
IDPLTHKSLLDGAGKSSDHSAPKSSVHDDKDDQNSNNKLSGSSSARFLNVRVANRFGH
RINHNVLSDIIGSNGLLTSHTPTTSVSEGERSTSSSSTHTSSNLPINRSITVDATSLSS
STFSDSPDPCLYEEIVGDIEDMTRFSSRCLSHVLSHEDLLMSVESCLENTSFMREITMIF
QEDKIETTSFNDSYVTPINEVDDSCGIDNYFG*
>G2346 (1..1011)
ATGGAGTTGTTAATGTGTTCCGGGTCAGGCCGAGTCAGGTGGTTCTTCTTCCACCGAGTCT
TCTTCACTCAGTGGTGGACTCAGGTTTGGTCAGAAGATCTACTTCGAGGATGGATCCGGA
TCCAGAAGCAAGAACCGGTCATACCGTTCCGTAAGTCGTCTACCACGGCGAGGTGCCAA
GTGGAAGGTTGTAGAAATGGATCTAAGCAATGTTAAAGCTTATTACTCGAGACACAAAGTT
TGTTGCATTCACCTCTAAATCATCTAAAGTCATTGTCTCTGCTCTTCAAAAGGTTTGT
CAACAATGTAGCAGGTTTACCAGCTTTCTGAGTTTGAAGTTGAGAGAAAAGAGTTGTGCG
AGAAGACTCGCTTGTCTAACAACGACGAAGAAAACCAACCCCAACACGGCTCTTTTC
ACTTCTCATTACTCTCGAATCGCTCCATCTCTTTACGGAACCCCAATGCTGCAATGATT

AAAAGCGTTTTGGGAGATCCTACTGCGTGGTCAACCGCAAGATCAGTGATGCAGCGGCCCT
GGACCGTGGCAGATTAATCCAGTTAGGAAAACCCATCCACACATGAATGTTTTATCACAT
GGAAGCTCAAGCTTTACTACATGTCCAGAGATGATAAACACAATAGCACAGATTCAAGC
TGTGCTCTCTCTCTCTGTCAAACCTACATCCCAATTCATCAGCAGCAACTTCAGACACCA
ACAAATACATGGCGACCATCTTCTGGTTTCGACTCGATGATCTCATTCTCCGATAAGGTT
ACAATGGCTCAGCCACCGCCCATTTCAACCCATCAGCCGCCCATCTCAACACATCAGCAG
TACCTCAGCCAACTTGGGAAGTCATCGCGGGCGAAAAGAGCAATTCACATTATATGTCT
CCTGTGAGTCAAATCTCGGAGCCAGCAGATTTCCAGATAAGCAATGGCAGTGTGTGCGCC
TATTCTCCTCCGTCCTTACTATCTCTTGTGTGCTACTTGCAGGCCCTATAG

>G2346 Amino Acid Sequence (domain in AA coordinates: 59-135)
MELLMCSGQAESGSSSTESSLSGGLRFGQKIYFEDGSGSRSKNRVNTVRKSSSTARCO
VEGCRMDLSNVKAYYSRHKVCCIHSKSSKIVISGLHQRFCQQCSRHFQLSEFDLEKRSCR
RRLACHNERRRRKPQPTTALFTSHYSRIAPSLYGNPNAAMIKSVLGDPTAWSTARSVMQRP
GPWQINPVRETHPHMNVLSHGSSSFTTCEMINNNSTDSSCALSLLSNSYPHQQQLQTP
TNTWRPSSGFDSMISFSDKVTMAQPPPISTHQPPISTHQYLSQTWEVIAGEKSNSHYMS
PVSQISEPADFQISNGSVSPSPSLLSLVCYLRLP*

>G237 (1..852)
ATGGCGAAGACGAAATATGGAGAGAGACATAGGAAAGGGTTATGGTCACCTGAAGAAGAC
GAGAAGCTAAGGAGCTTCATCCTCTCTTATGGCCATTCTTGCTGGACCACTGTTCCCATC
AAAGCTGGGTTACAAAGGAATGGGAAGAGCTGCAGATTAAGATGGATTAAATTACCTAAGA
CCAGGGTTAAAGAGGGATATGATTAGTGCAGAAGAAGAAGAGACTATCTTGACGTTTCAT
TCTCCCTTGGGTAAACAAGTGGTCGCAATAGCTAAATTCTTACCGGAAGAACAGACAAT
GAGATAAAGAACTATTGGCACTCTCATTGAAAAAGAAATGGCTCAAGTCTCAGAGCTTA
CAAGATGCAAAATCTATTTCCCTCTCTCGTCTTCATCATCACTTGTGTGCTGTGGA
GAAAGAAATCCGGAACCTTGATCTCGAATCACGTGTTCTCCCTCCAGAGACTTCTAGAG
AACAAATCTTCATCTCCCTCACAGAAAGCAACGGAAATAACAGCCATCAATGTTCTTCT
GCTCCTGAGATTCGAAGCTTTTCTCTCTGAATGGCTTTCTTCTTCATATCCCCACACC
GATTATTCCTCTGAGTTTACCGACTCTAAGCACAGTCAAGCTCCAAATGTCGAAGAGACT
CTCTCAGCTTATGAAGAAATGGGTGATGTTGATCAGTTCCATTACAACGAAATGATGATC
AACACAGCAACTGGACTCTTAACGACATTGTGTTTGGTTCCAAATGTAAGAAGCAGGAG
CATCATATTATAGAGAGGCTTCAGATTGTAATTCTTCTGCTGAATTCTTTCTCCACCA
ACAACGACGTAAATTGCGTTTATTGTAATGTAAATCAAATTTCTAAGGCAAAACCGGAAA
AAAAAAAAAAAAAAAAAAAA

>G237 Amino Acid Sequence (domain in AA coordinates: 11-113)
MAKTKYGERHRKRLWSPEEDEKLRSFILSYGHSCWTTVPKAGLQRNGKSCRLRWINYLR
PGLKRDMLISABEEETILTFHSPPLGNKWSQIAKFLPGRTDNEIKNYWHSLLKKKWLKSQSL
QDAKSISPPSSSSSSSLVACGERNPETLISNHVFSLQRLLENKSSSPSQESNGNNSHQCSS
APEIPLRFSEWLSSSYPHTDYSSEFTDSKHSQAPNVEETLSAYEEMGDVDQFHYNEMMI
NNSNWLNDIVFGSKCKKQEHHIYREASDCNSSAEFFSPPTTT*

>G2373 (48..1199)
GCAAAATCCTCAGATCGTCTTACCTTCTCCGAATCGATCGATTTTTCATGGAGGACGACG
ACGAGATTCAGTCAATTCATCTCCGGGAGATTCTTCCCTTTTACCACAAGCTCCTCCTT
CTCCGCCGATTTTGCCAAACAAACGACGTGACGGTGGCCGTCGTGAAGAAACCACAACCGG
GGCTTTCTTCTCAATCTCCGTCCATGAACGCTTTAGCGTTAGTGGTTCATACTCCTTCTG
TAACCGGTGGTGGTGGTAGCGGAAACAGAAACGGACGAGGAGGAGGAGGAAGCGGTG
GTGGTGGAGGAGGAAGAGATGATTGTTGGAGCGAAGAAGCTACAAAGGTTCTAATCGAAG
CTTGGGGAGATCGATTCTCTGAACAGGTAAAGGAACCTTTGAAGCAACAACATTGGAAAG
AAGTAGCTGAGATTGTGAACAAGAGTCGTCAATGCAATACCCATAAACTGATATTCACT
GTAAGAACAGAATTGATACGGTGAAGAAGAAGTATAAGCAAGAGAAAGCTAAGATTGCTT
CTGGTGATGGACCTAGTAAATGGGTTTTCTTCAAGAAGCTTGAGAGTTTGATTGGTGGTA
CTACAACATTTCATTGCTTCTTCAAAAGCTTCAGAGAAGGCTCCTATGGGAGGAGCTCTTG
GGAATAGCCGTTTCGAGTATGTTTAAACGGCAAACCTAAAGGTAATCAGATTGTGAGCAAC
AACAAAGAGAAGAGAGGCTCTGATTCGATGCGGTGGCATTTTAGGAAACGTTAGTGTCTG
AGACTGAGTCTGAGTCTGATCCTGAACCTGAGGCTTCTCCTGAGGAATCTGCTGAGAGTC
TCCCACCTTTGCAACCGATTCAACCGCTTTCGTTTCATATGCCAAAGCGGTTGAAGGTGG
ATAAGAGTGGAGGTGGAGGAGTGGAGTGGAGATGTGGCGAGGGCGATACTTGGATTTA
CGGAAGCTTATGAGAAGCGGAAACTGCTAAGCTTAAGTTAATGGCGGAACCTGGAAAAGG

AGAGGATGAAATTTGCTAAAGAGATGGAGTTGCAGAGAATGCAGTTCTTGAAAACCTCAAT
TGGAGATAACACAGAACAAATCAAGAAGAGGAAGAGAGGAGCAGGCAGCGAGGAGAAAGGA
GGATCGTTGATGATGATGATGATCGCAATGGCAAGAATAACGGCAATGTAAGTAGCTGAC
AATTGAACACACAAATGTTCTATGATATTTGCTATGATAAGCTGGATTTTAGGTTTTGA
TGG

>G2373 Amino Acid Sequence (domain in AA coordinates:290-350)
MEDDDDEIQSIPSPGDSSLSPPQAPPSPPIPTNDVTVAVVKKPQPGLSSQSPSMNALALVV
HTPSVTGGGSGNRNGRGGGGSGGGGGRDDCWSEETKVLIEAWGDRFSEPGKGTLLKQ
QHWKEVAEIVNKSQCKYPKTDIQCKNRIDTVKKKYKQEKAKIASGDGPSKWVFFKKLES
LIGTTTTFIASSKASEKAPMGGALGNSRSSMFKRQTKGNQIVQQQEKRGSDSMRWHFRK
RSASETESESDPEPEASPEESAESLPPLQPIQPLSFHMPKRLKVDKSGGGSGVGDVARA
ILGFTEAYEKAETAKLKLMAELEKERMKFAKEMELQRMQFLKTQLEITQNNQEEEEERSRQ
RGERRIVDDDDDRNGKNNGNVSS*

>G2376 (39..1370)

CACGAGCTTCGACTCAGATCCGGCGATATCGAATTCATGGAGGACGATGAAGACATCC
GATCTCAGGGTTCCGATTACCTGATCCGTCTTCTCCCCCGCGGGGACGAATCACGG
TTACGGTGGCTTCGGCAGGTCCGCCCTTCTTATTCTCTGACTCCTCCGGGTAATTCGTCCG
AGAAGGATCCGGATGCGTTGGCTCTGGCGCTGCTTCCGATTACAGCCAGCGGTGGAGGGA
ATAACAGCAGTGGGAGACCAACCGGCGCGCGGGAGGAGATTGTTGGAGCGAAGCAG
CTACGGCTGTGTTGATTGATGCGTGGGGTGAGAGATACTTGGAGCTTAGCAGAGGGAATC
TGAAGCAGAAGCACTGGAAGAGGTGGCTGAGATTGTGAGCAGCAGAGAGGATTACGGTA
AAATTTCCCAAACCTGATATACAGTGTAAAGATAGGATCGATACGGTGAAGAAGAAGTATA
ACAAGAGAAGGTGAGAATCGCTAACGGCGGTGGCCGTAGCAGATGGGTGTTCTTCGACA
AGCTTGACCGTCTGATTGGATCAACGGCGAAGATCCCGACGGCAACTTCTGGAGTCAGCG
GTCTCTGTCGGAGGATTGCATAAGATTCTATGGGTATTCCAATGGGAAGTCGTTTCAATC
TGTACCATCAGCAAGCTAAGGCTGCAACACCGCCTTTCAATAATCTTGACCGGTTAATTG
GAGCTACGGCTAGAGTCTCAGCTGCTTCTTTCGGTGGCAGTGGTGGAGGAGGCGGAGGAG
GATCTGTCAATGTACCTATGGGAATTCGGATGAGTAGCCGTTTCACTCCGTTTGGACAGC
AAGGGAGGACTCTGCCACAGCAAGGTAGGACACTGCCACAGCAACAGCAGCAAGGGATGA
TGGTGAAGAGGTGTAGTGAGTCGAAACGCTGGCGTTTCAGGAAGAGGAACGCTTCTGATT
CAGACTCGGAATCTGAAGCAGCAATGTCAGATGATTCCGGTGACAGTTTACCACCTCCTC
CTCTGTGCGAAGAGGATGAAGACGGAGGAGAAGAAGAAGCAAGATGGTGTATGGAGTGGGGA
ACAAATGGAGGGAGCTGACTCGGGCAATCATGAGATTCCGTGAAGCTTATGAGCAAACAG
AGAATGCGAACTGCAACAGGTGGTTGAGATGGAGAAAGAGAGGATGAAGTTCTTGAAGG
AGCTTGAGTTGCAGAGAAATGCAGTTCTTTGTGAAGACTCAATTGGAGATATCACAACCTTA
AGCAGCAACATGGGAGGAGAAATGGGAAACACCAGTAATGATCATCATCACAGCCGCAAGA
ACAACATCAATGCGATTGTCAACAACAACAGATTGGGTAATAACTAGAATTTAGTGA
TGCAGTGTGTAATTGATATATTTTAGATTTGAG

>G2376 Amino Acid Sequence (domain in AA coordinates:79-178, 336-408)

MEDDEDIRSQSDSPDPSSSPAGRITVTVASAGPPSYSLTPPGNSSQKDPDALALALLP
IQASGGGNSSGRPTGGGGREDWCSEAAAVLIDAWGERYLELSRGNLQKQHWKEVAEIV
SSREDYGKIPKTDIQCKNRIDTVKKKYKQEKVRIANGGGRSRWVFFDKLDRLLIGSTAKIP
TATSGVSGPVGGLHKIPMGIPMGSRSNLYHQAKAATPPFNNLDRLLIGATARVSAASFGG
SGGGGGGGSVNVPMPGIPMSSRSAPFGQQGRTLPPQQGRTLPPQQQQGMMVKRCSESKRWF
RKRNASDSDSESEAAAMSDSGDSLPPPPPLSKRMKTEKKKQDGDGVGNKWRELTRAIMRF
GEAYEQTENAKLQVVEMEKERMKFLKELELQRMQFFVKTQLEISQLKQHGRRMGNTSN
DHHHSRKNINAIIVNNNDLGN*

>G24 (194..724)

CGGACGCGTGGGCAAATATTAAAATAAAAAGTGTCCGGTGAATTCTCAATCTTTGTCTTCT
TTCGTCTCTCTTTAAACTCCTCCGTCCCTCCTTATTATGTAACCGTCTCGCCGTCAA
TTTCAAAATCTCTCCCTCCGTTCATAAACCCAGATCGAAATTTATGGTTTTGTAAATTT
TTTACCGCGGTTATGGAGACGGAAGAGAGATCTGAAACCGTATAAAGGAATACGAATGAGGAAATG
GGGGATTGGGACGAGGAAGAGAGATCTGAAACCGTATAAAGGAATACGAATGAGGAAATG
GGGGAATGGGTGGCGGAGATACGGGAACCGAATAAGAGATCAAGGATCTGGTTAGGTTT
TTATGCGACGCTGAAGCGGCGGCGAGAGCTTACGACACTGCTGTTTTTTTACCTCCGTGG
TCCTTCAGCGAGGCTTAATTTTCCGGAGCTTTTGGCTGGACTTACTGTTTTTCAACGGCGG
AGGAAGAGGTGGTGAATTTATCGGCGGCGTATATTAGGAGAAAAGCGGCGGAGGTGGTGC

TCAGGTTGATGCGCTTGGAGCGACGGTGGTGTGAATACCGGCGGCGAGAATCGCGGTGA
TTACGAGAAGATTGAGAATTGTCGTAAGAGCGGTAAACGGGTCATTGGAACGGGTCGATTT
GAATAAATTACCCGACCCGAAAATTCCGATGGTGATGATGACGAATGTGTGAAAAGAAG
ATAGAAAAAATAAAAAGTAGTTGTAGAAGGAGAGACGAGAATGTTTGTCTTTAAGATGCG
CTGTTGCCGCTAACATGCGCTTTCGATTTTAGTGTAAACATGCGCCTCCATTGTTTTTG
GGTTTTGTTTTCGTCGTCGATAATCAAAGATTTTAAACACAATTCTCAAATTTTCACT
TGTTACAACTAGATTTGCATGATCTTTGTATTAACGAATAACGATTAAGTCCTAAA
>G24 Amino Acid Sequence (domain in AA coordinates: 25-93)
METEAAVTATVTAATMGIGTRKDLKPYKGIRMRKWGWAEIREPNKRSRIWLGSYATP
EAAARAYDTAVFYLRGPSARLNFPELLAGLTVSNGGGRGGLSAAIYIRKAAEVGAQVDA
LGATVVVNTGGENRGDYEKIENCRKSGNSLERVDLNLKLPDENSDGDDDECVKRR*
>G2424 (1..999)
ATGAGGATGGAGATGGTGCATGCTGACGTGGCGTCTCTCTCCATAACACCTTGCTTCCCG
TCTTCTTTGTCTTCGTCTCATCATCACTATAACCAACAACAACATTGTATCATGTGCG
GAAGATCAACACCATTCGATGGATCAGACCACTTCATCGGACTACTTCTCTTTAAATATC
GACAATGCTCAACATCTCCGTAGCTACTACACAAGTCATAGAGAAGAAGACATGAACCTT
AATCTAAGTGATTACAGTAATTGCAACAAGAAAGACACAACAGTCTATAGAAGCTGTGGA
CACTCGTCAAAAGCTTCGGTGTCTAGAGGACATTGGAGACCAGCTGAAGATACTAAGCTC
AAAGAACTAGTCGCCGTCTACGGTCCACAAAACCTGGAACCTCATAGCTGAGAAGCTCCAA
GGAAGATCCGGGAAAAGCTGTAGGCTTCGATGGTTTAAACCACTAGACCCAAGGATAAAT
AGAAGAGCCTTCACTGAGGAAGAAGAAGAGAGGCTAATGCAAGCTCATAGGCTTTATGGT
AACAAATGGGCGATGATAGCGAGGCTTTCCCTGGTAGGACTGATAATTCTGTGAAGAAC
CATTTGGCATGTTTATAATGGCTCGCAAGTTAGGGAACAATCTTCTTCTTACCGTAGGAGG
AAGACGATGGTTTCTCTTAAGCCACTCATTAAACCCTAATCCTCACATTTTCAATGATTTT
GACCCTACCCGGTTAGCTTTGACCCACCTTGCTAGTAGTGACCATAAGCAGCTTATGTTA
CCAGTTCCTTGCTTCCAGGTTATGATCATGAAAATGAGAGTCCATTAATGGTGGATATG
TTCGAAACCCAAATGATGGTTGGCGATTACATTGCATGGACACAAGAGGCAACTACATTC
GATTTCTTAAACCAAACCGGGAAGAGTGAGATATTTGAAAGAATCAATGAGGAGAAGAAA
CCACCATTTTTTCGATTTTCTTGGGTTGGGGACGGTGTGA
>G2424 Amino Acid Sequence (conserved domain in AA coordinates: 107-219)
MRMEMVHADVASLSITPCFPSSLSSSSHHYNNQQQHCHIMSEDQHHSMDQTTSSDYFSLNI
DNAQHLRSYYTSHREEDMNPNSLDYSNCKKDTTVYRSCGHSSKASVSRGHWPAEDTKL
KELVAVYGPQNWNLIAEKLQGRSGKSCRLRWFNQLDPRINRRAFTEEEERLMQAHRLYG
NKWAMIARLFPGRDTSVKNHWHVIMARKFREQSSSYRRRKTMSVLKPLINPNPHIFNDF
DPTRLALTHLASSDHKQLMLPVPCFPGYDHENESPLMVDMEFETQMMVGDYIAWTQEATTF
DFLNQTKKSEIFERINEEKKPPFFDFLGLGTV*
>G2505 (1..1026)
ATGGGTTCTTCGTGCAACGGAGGAGTGCCACCTGGTTTCCGGTTTCATCCGACGGACGAA
GAGCTTCTCCATTACTACTTGAAGAAGAAAATCTCTTACCAAAAAGTTTGAGATGGAAGTC
ATCAGAGAGGTTGACTTAAACAAGCTTGAGCCTTGGGATTTGCAAGAGAGATGCAAGATA
GGATCAACACCACAACGAATGGTACTTCTTCAGCCACAAGGACAGGAAATATCCGACG
GGGTCAAGGACCAACCGTGCTACTCATGCAGGGTTCTGGAAGGCGACGGGACGTGACAAG
TGCATAAGGAACTCTTACAAAAGATAGGAATGAGGAAGACACTTGTGTCTACAAAGGT
AGAGCTCCTCATGGCCAAAAGACTGATTGGATCATGCATGAGTACCGTCTTGAAGACGCT
GATGATCCTCAAGCCAACCTAGTGAAGATGGATGGGTGGTATGTAGAGTGTATGAAG
AAAAATTTGTTCAAGGTAGTAAATGAAGGTAGCTCAAGCATTAACTATTGGACCAACAC
AACCATGACGCATCTAACAACAACCATGCACCTCAAGCTCGTAGCTTTATGCACGAGAC
AGTCCATACAGCTAGTACGTAAACCACGGAGCCATGACATTGCAACTTAACAAGCCTGAC
CTTGCTCTTCATCAATACCCACCAATCTTCCACAAGCCACCTTCACTTGGATTTGACTAC
TCTTCAGGACTTGCAAGGGACAGTGAGAGTGGCGCTAGTGAAGGGTTACAATACCAGCAA
GCGTGTGAGCCGGGTTTAGACGTTGGTACATGTGAGACAGTGGCTAGTCATAATCATCAA
CAAGGTCTAGGTGAATGGGCAATGATGGATAGACTTGTGACTTGTACATGGGAAATGAA
GATTCCTCTAGAGGGATTACGTATGAGGATGGTAACAACAATTTCGTCTCTGTGGTTCAG
CCAGTTCGCCGACGAACCAGCTAACATTGCGTAGTGAGATGGATTTCTGGGGTTATTCT
AAATAG
>G2505 Amino Acid Sequence (domain in AA coordinates: 10-159)
MGSSSNGGVPPGFRFHPTDEELLHYLKKKISYQKFEMEVIREVLDLNLKLEPWDLQERCKI

GSTPQNEWYFFSHKDRKYPTGSRTNRATHAGFWKATGRDKCIRNSYKKIGMRKTLVIFYKG
RAPHGQKTDWIMHEYRLEDADDPQANPSEDGWVVCVRFVFMKKNLFKVVNEGSSSINSLDQH
NHDASNNNHALQARSFMHRDSPYQLVRNHGAMTFELNKPDLALHQYPPIFHKPPSLGFDY
SSGLARDSESAASEGLQYQACEPGLDVGTCETVASHNHQQGLGEWAMMDRLVTCHEMGNE
DSSRGITYEDGNNSSSSVVQPVPA TNQLTLRSEMDFWGYSK*

>G2512 (64..798)

AACTTAGTGCCACTTAGACACAATAAGAAAACCGTTAACAAGAAGAAAAAAGATCG
AAAATGGAATATCAAACTAACTTCTTAAGTGGAGAGTTTTCCCGGAGAACTCTTCTTCA
AGCTCATGGAGCTCACAGAATCATTCTTGTGGGAAGAGAGTTTCTTACATCAATCATT
GACCAATCCTTCCCTTTTATCTAGCCCTACTGATACTACTGTGATGACTTCTTGCATT
GAATCATCAATCATAAAAGAAGAAGGAAAAGAAGCCACCGTGGCGGCCGAGGAGGAGGAG
AAGTCATACAGAGGAGTGAGGAAACGGCCGTGGGGGAAATTCGCGGCCGAGATAAGAGAC
TCAACGAGGAAAGGGATAAGAGTGTGGCTTGGGACATTCGACACCGCGGAGGCGCGGCT
CTCGCTTATGATCAGGCGGCTTTCGCTTTGAAAGGCAGCCTCGCAGTACTCAATTTCCCG
GCGGATGTCTGTTGAAGAACTCTCTCCGGAAGATGGAGAATGTGAATCTCAATGATGGAGAG
TCTCCGGTGATAGCCTTGAAGAGAAAACACTCCATGAGAAACCGTCTAGAGGAAAGAAG
AAATCTTCTTCTTCTTCGACGTTGACATCTTCTCCTTCTTCTCCTCCTCTATTCTCT
TCTTCGTCTTCTTCTTCTTCTGTCGTCAGAAAGTAGAAAACAGAGTGTGTTATGACGCAA
GAAAGTAATACAACACTTGTGGTCTTGTAGGATTTAGGTGCTGAATACTTAGAAGAGCTT
ATGAGATCATGTTCTTGATAATCTCTGCTTCTACAATTTTTATGTAATTGA

>G2512 Amino Acid Sequence (conserved domain in AA coordinates: 79-139)

MEYQTNFLSGEFSPENSSSSSSWSSESFLWEESEFLHQSFDQSFLSSPTDNYCDDFFAFE
SSIIKEEGKEATVAEEEEESYRGVVRKRPWGKFAAEIRDSTRKGI RVWLGTDFDTAEAAAL
AYDQAAAFALKGSLAVLNFPADVVEESLRKMENVNLDGESPVIALKRKHSRNRPRGKKK
SSSSSTLTSSPSSSSSYSSSSSSSLSSRSRKQSVVMTQESNTTLVVLEDLGAELYLEELM
RSCS*

>G2513 (69..698)

TTTCAACAGTAATTTAAGTTAACCGGAGTCTCTTTTGTGTTTTCCGGCGAATTTTTGGTAC
TTTGAGTTATGAATAATGATGATATTATTCTGGCGGAGATGAGGCCTAAGAAGCGTGCGG
GAAGGAGAGTGTTTAAAGGAGACACGTCACCCAGTTTACAGAGGCATAAGGCGGAGGAACG
GTGACAAATGGGTCTGCGAAGTCAGAGAACCGACGCACCAACGCCGCAATTTGGCTCGGGA
CTTATCCACAGCAGATATGGCAGCGCGTGACACGACGTGGCGGTTTTAGCTCTGCGTG
GGAGATCCGCATGTTTGAATTTCCGCCGACTCCGCTTGGCGGCTTCCGGTGCCGGAATCCA
ATGATCCGGATGTGATAAGAAGAGTTGCGGCGGAAGCTGCGGAGATGTTTAGGCCGGTG
ATTTAGAAAGTGAATTACGGTTTTGCCTTGTGCGGAGATGATGTGGATTGGGTTTTG
GTTCCGGTTCCCGCTCTGGTTCCGGATCGGAGGAGAGGAATCTTCTTCGTATGGATTG
GAGACTACGAAGAAGTCTCAACGACGATGATGAGACTCGCGGAGGGGCCACTAATGTCCG
CGCCGCGATCGTATATGGAAGACATGACTCCTACTAATGTTTACACGGAAGAAGAGATGT
GTTATGAAGATATGTCATTGTGGAGTTACAGATATTAAGTGGGACTCACATATCTACTAT
ACATAATATTAGCTTTTATGTAAAGGTTATTTATGTGAGTTTTAAGATTGTAGATGTGT
CCAGGCGTTAGAAGTTTCTTGTATGGTATGGAATCTTTGTACCTATAAAATTATAAAAT
T

>G2513 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNDIILAEMRPKKRAGRRVFKETRHPVYRGIRRRNGDKWVCEVREPTHQRRRIWLGTY
TADMAARAHDVAVLALRGRSACLNFADSAWRLPVPESNDPDVIRRVAAEAEMFRPVDLE
SGITVLPAGDDVDLGFSGSGSGSGSEERNSSSYGFGDYEEVSTTMMRLAEGPLMSPPR
SYMEDMTPTNVYTEEMCYEDMSLWSYRY*

>G2519 (83..691)

CAAAGTGAAAACATAAGATCATCTTCTTCGTTGATAGATCAATATAGGAACTCCAGAAGA
GAATCTTGATCAATTAAGTATCATGTCTCACATCGCTGTTGAAAGGAATCGAAGAAGGCA
AATGAACGAGCATCTTAAATCCCTTCGTTCTTTGACTCCTTGTCTTACATCAAAGGGG
AGATCAAGCTTCGATCATCGGAGGAGTGATAGAGTTTCAAAAGAGTTGCAGCAATTGGT
TCAAGTTCTTGAGTCCAAGAAACGTCGAAAGACCCTAAACCGACCATCTTCCCTTATGA
TCACCAGACAATCGAGCCATCCAGTTTAGGAGCCGCCACTACCCGAGTACCGTTTAGTCG
AATCGAAAATGTGATGACCACAAGTACTTTCAAGGAAGTAGGAGCATGCTGTAACCTCCC
TCATGCTAACGTAGAAGCAAAGATTTCAAGTTCTAATGTTGTATTGAGAGTTGTCTCTAG
CGGAATCGTGGGCGAGCTCGTAAAGATCATCTCTGTCTTAGAGAAGCTATCTTTCAAGT

TCTTCACCTCAATATTAGTAGCATGGAGGAGACTGTCTTATACTTTTTCGTTGTTAAGAT
AGGATTGGAGTGTCACTTAAGCTTGGAGGAGCTAACTCTTGAAGTTCAGAAAAGCTTTGT
GTCTGATGAAGTGATCGTCTCTACCAATTA AAAACAAAATCTACATGTACTAGAGCGTG
TATCGTTTTTTGGGATTAATAATCATATAATCGTTACATGAGCCTTGATACTTTGCTAGA
AATAAGCTCCTCTAAACAAAACCTTCTTTTTAAAAAACACACTTATGTTTTACTTAGTT
TGTTGTTGTATCCGAAGTTGATCAACGTTGTAATTTCCCAATAAATCATGACATTTTA
TATGCTCT

>G2519 Amino Acid Sequence (domain in AA coordinates:1-65)

MSHIAVERNRRRQMEHLKSLRSLTPCFYIKRGDQASIIGGVIEFIKELQQLVQVLESKK
RRKTLNRPSFPYDHTIEPSSLGAATTRVPFSRIENVMTTSTFKEVGACCNSPHANVEAK
ISGSNVVLRVVSRRIVQLVKIISVLEKLSFQVLHLNISSMEETVLYFFVVKIGLECHLS
LEELTLEVQKSFVSDEVIVSTN*

>G2520 (133..1197)

AAGGAGTTTTGCATACTACCAAGCCACAATCATTTCTCTCTCTATCTCTCTGGTTT
TGAATCGGCGACGACTGAGTCAACTCGGTGTTGTTACTGGTTTCGTCGTATGTGTTGTAA
CTGATTAAGTTGATGGATCCGAGTGGGATGATGAACGAAGGAGGACCGTTTAATCTAGCG
GAGATCTGGCAGTTTCCGTTGAACGGAGTTTCAACCGCCGGAGATTCTTCTAGAAGAAGC
TTCGTTGGACCGAATCAGTTCGGTGATGCTGATCTAACCACAGCTGCTAACGGTGATCCA
GCGCGTATGAGTCACGCGTTGTCTCAGGCGGTTATTGAAGGTATCTCCGGCGCTTGGAAA
CGGAGGGAAGATGAGTCTAAGTCGGCGAAGATCGTCTCCACCATTGGCGCTAGTGAAGGT
GAGAACAAAAGACAGAAGATAGATGAAGTGTGTGATGGGAAAGCAGAAGCAGAATCGCTA
GGAACAGAGACGGAACAAAAGAAGCAACAGATGGAACCAACGAAAGATTATATTCATGTT
CGAGCTAGAAGAGGTCAAGCTACTGATAGTCACAGTTTAGCTGAAAGAGCGAGAAGAGAG
AAAATAAGTGAGCGGATGAAAATCTTGCAAGATCTTGTTCCGGGATGTAACAAGGTTATT
GGAAAAGCACTTGTTCTAGATGAGATAATTAATATATACAATCATTGCAACGTCAAGTT
GAGTCTTATCGATGAAGCTTGAAGCAGTCAACTCAAGAATGAACCCTGGTATCGAGGTT
TTTCCACCCAAAGAGGTGATGATTCTCATGATCATCAACTCAATCTTCTCCATTTTTTTC
ACAAAACAATACATGTTTCTATCGAGGTATTCTCGGGGTAGGAGTCTCGATGTTTATGCG
GTTTCGGTCATTTAAGCATTGCAATAAACGGAGTGACCTCTGTTTTTGTCTCTGCTCCCCA
AAAACAGAACTTAAGACAATATATTTTACAAAACATGACATGTTTCTGTGATATTCT
CGAGTAGGAGTCGCTATTAGTTCATCTAAGCATTGCAATGAACCGTTTGGTCAGCAAGCG
TTTGAGAATCCGGAGATACAGTTCGGGTCGCAGTCTACGAGGGAATACAGTAGAGGAGCA
TCACCAGAGTGTTGTCACATGCAGATAGGATCAGGTGGTTTCGAAAGAACGTCTTGA

>G2520 Amino Acid Sequence (domain in AA coordinates: 135-206)

MDPSGMMNEGGPFNLAEIWQFPLNGVSTAGDSSRRSFVGPNOFGDADLTAAANGDPARMS
HALSQAVIEGISGAWKRREDESKSAKIVSTIGASEGENKRQKIDEVCDGKAEBSLGTET
EQKKQQMEPTKDYIHVRARRQATDSHSLAERARREKISERMKILQDLVPGCNKVIKAL
VLDEIINYIQLQRQVEFLSMKLEAVNSRMNPGIEVFPKPEVMILMIINSIFSIFFTKQY
MFLSRYSRGRSLDVYAVRSFKHCNKRSDLCFCSCSPKTELKTTIFSQNMTCFCRYSRVGV
AISSSKHCFNEFPFQAFENPEIQFGSQSTREYSRGASPEWLHMQIGSGGFERTS*

>G2533 (1..1080)

ATGATAAGCAAGGATCCAATATCGAGTTTACCTCCAGGGTTTCGATTTTCATCCAACAGAT
GAAGAACTCATTCTCCATTACCTAAGGAAGAAAGTTTCTCTTCCCCAGTCCCGCTTTCG
ATTATCGCCGATGTGATATCTACAAATCCGATCCATGGGATTTACCAGCTAAGGCTCCA
TTTGGGGAGAAAGAGTGGTATTTTTCAGTCCGAGGGATAGGAAATATCCAAACGGAGCA
AGACCAAACAGAGCAGCTGCGTCTGGATATTGGAAGCAACCGGAACAGATAAATTGATT
GCGGTACCAAAATGGTGAAGGGTTTCATGAAAACATTGGTATAAAAAAGCTCTTGTTGTTT
TATAGAGGAAAGCCTCCAAAAGGTGTTAAACCAATTGGATCATGCATGAATATCGTCTT
GCCGATTTCATTATCTCCAAAAGAAATTAACCTCTTCTAGGAGCGGTGGTAGCGAAGTTAAT
AATAATTTTGGAGATAGGAATCTAAAGAAATATTGATGAGACTGGATGATTGGGTCTCTT
TGCCGGATTTACAAGAAATCACACGCTTCATTGTATCACCTGATGTTGCTTTGGTCACA
AGCAATCAAGAGCATGAGGAAATGACAACGAACCATTCGTAGACCGCGGAACCTTTTTCG
CCAAATTTGCAAAATGATCAACCCCTTAAACGCCAGAAGTCTTCTTGTTTCGTTCTCAAAC
TTACTAGACGCTACAGATTTGACGTTTCTCGCAAATTTCTAAACGAAACCCCGGAAAAT
CGTTCTGAATCAGATTTTCTTTTCATGATTGGCAATTTCTCTAATCCTGACATTTACGGA
AACCATTACTTGGATCAGAAGTTACCGCAGTTGAGCTCTCCCACTTCAGAGACAAGCGGC
ATCGGAAGCAAAAGAGAGAGAGTGGATTTTGCAGGAAGAAACGATAAACGCTTCGAAGAAG

ATGATGAACACATATAGTTACAATAATAGTATAGATCAAATGGATCATAGTATGATGCAA
 CAACCTAGTTTCCCTGAACCAGGAACCTCATGATGAGTTCTCACCTTCAATATCAAGGCTAG
 >G2533 Amino Acid Sequence (conserved domain in AA coordinates:11-186)
 MISKDPISSLPPGFRFHPTDEELILHYLRKKVSSSPVPLSIIADVDIYKSDPWDLPAKAP
 FGEKEWYFFSPDRKYPNGARPNRAASGYWKATGTDKLIAPNPEGFHENIGIKKALVF
 YRGKPPKGVKTNWIMHEYRLADSLSPKRINSSRSGGSEVNNFNGDRNSKEYSMRLDDWVL
 CRIYKKSHASLSSPDVALVTSNQEHEENDNEPFVDRGTFPLNLQNDQPLKRQKSSCSFSN
 LLDATDLTFLANFLNETPENRSESDFSFMIGNFSNPDYGNHYLDQKLPQLSSPTSETSG
 IGSKRERVDFAETINASKKMMNTYSYNNSIDQMDHSMMQPSFLNQELMMSSHLQYQG*
 >G2534 (1..975)

ATGGATAATATAATGCAATCGTCAATGCCACCGGGATTCCGATTTCATCCGACAGAGGAA
 GAGCTTGTGGGTTATTACCTAGATAGGAAGATCAATTCAATGAAGAGTGCTTTAGATGTC
 ATTGTAGAGATTGATCTCTACAAAATGGAGCCATGGGATATACAAGCGAGGTGTAAACTA
 GGGTATGAAGAGCAAAACGAGTGGTACTTCTTTAGTCATAAGGACAGGAAGTACCCTACC
 GGGACTAGGACCAACCGAGCCACTGCGGCTGGGTTCTGGAAAGCCACGGGTAGAGACAAG
 GCGGTACTATCAAAAAACAGTGTGCATCGGAATGCGGAAGACACTTGTCTACTACAAGGGT
 CGAGCTCCTAATGGAAGAAAGTCCGATTGGATCATGCACGAATACCGTCTCCAAACTCC
 GAGCTTGCCCCGGTTCAGGAGGAAGGCTGGGTGGTGTGTGCGAGCATTTAGGAAGCCAATT
 CCAAACAGAGGCCATTAGGGTACGAGCCATGGCAGAACCAGCTCTACCACGTCGAAAGT
 AGTAACAACACTACTCATCTTCAGTGACAATGAACACGAGTCATCATATCGGTGCATCTTCA
 TCAAGTCATAACCTTAATCAAATGCTCATGAGCAATAACCACTACAATCCTAATAATACA
 TCCTCATCGATGCATCAATATGGCAACATTGAGCTCCCGCAGTTGGACAGCCCCGAGCTTG
 TCGCCTAGTTTAGGACGAATAAAGATCAGAACGAGAGTTTCGAGCAAGAAGAAGAGAAG
 AGCTTTAACTGTGTGGATTGGAGAACACTAGATACCTTGCTTGAGACACAAGTCATACAT
 CCGCATAACCCTAATATTCTTATGTTGAAACGCAGTCGTATAATCCGGCGCCAGCTTC
 CTTTCCATGCATCAAAGCTATAATGAGGTCGAAGCTAATATTCATCATTCTCTTGATGC
 TTCCCTGACTCGTAA

>G2534 Amino Acid Sequence (conserved domain in AA coordinates:10-157)
 MDNIMQSSMPPGFRFHPTDEELVGYLDKINSMSALDVIVEIDLYKMEPWDIQRCKL
 GYEEQNEWYFFSHKDRKYPTGTRTNRATAAGFWKATGRDKAVLSKNSVIGMRKTLVYYKG
 RAPNGRKS DWIMHEYRLQNSLAPVQEEGWVVCRAFRKPIPNQRPLGYEPWQNLHYHVES
 SNNYSSSVTMTNTHHIGASSSSHNLMMSNNHYNPNNTSSSMHQYGNIELPQLDSPSL
 SPSLGTNKDQNESFEQEEESFNCVDWRTLDTLLETQVIHPHNPNI LMFETQSYPAPSF
 PSMHQSYNEVEANIHHS LGCFPDS*

>G2573 (34..957)
 CCAGATTTAATTGAGACTCTCAAAGAAACACCATGGAAGAAGAGCAACCTCCGGCCAAG
 AAACGAAACATGGGGAGATCTAGAAAAGGTTGCATGAAAGGTAAAGGCGGTCCAGAGAAC
 GCCACGTGTACTTTCCGTGGAGTTAGGCAACGGACTTGGGGTAAATGGGTGGCTGAGATC
 CGTGAGCCTAACCGTGGGACTCGTCTCTGGCTCGGCACGTTTAATACCTCGGTCCGAGGCC
 GCCATGGCTTACGATGAAGCCGCTAAGAACTCTATGGACACGAGGCTAAACTCAACTTG
 GTGCACCCACAACAACAACAAGTAGTAGTGAACAGAACTTGTCTTTTCTGGCCAC
 GGGTCGGGTTCTTGGGCTTATAATAAGAAGCTCGATATGGTTCATGGGTGGACCTTGGT
 CTCGGCCAGGCAAGTTGTTACAGAGGTTCTTGCTCAGAGAGATCGAGTTTTCTACAAGAA
 GATGATGATCATAGTCATAATCGATGTTTCGTCTTCAAGTGGTTTCAATCTTTGTTGGTTA
 TTACCTAAACAAAGTGATTACAAAGATCAAGAGACCGTTAATGCTACGACTAGTTATGGC
 GGTGAAGGCGGTGGTGGCTTACGTTAACGTTTTTCGACCAATTTGAAACCAAAGAATTTG
 ATGAGTCAGAATTATGATTATACAATGGAGCTTGGTCTAGGTTTTCTGTGGGGCAAGAA
 AAGAAGACGGAACATGACGTGTCTCGTGTGGATCGTCGGACAACAAGGAGAGTATG
 TTGGTTCCTAGTTGCGGCGGAGAGAGGATGCATAGCCGGAGTTGGAAGAGCGAACAGGA
 TATTTGGAATGGATGATCTTTTGGAGATTGATGATTAGGTTTGTGATTGGCAAAAAT
 GGAGATTTCAAGAATTGGTGTGTGAAGAGTTTCAACATCCATGGAATTGGTCTGAGAG
 TTTTATTTATTACTATTATTTATCATACATATTTCTTATATTTGACTTAGG

>G2573 Amino Acid Sequence (domain in AA coordinates: TBD)
 MEEEQPPAKKRNMGSRKGMKGKGGPENATCTFRGVQRQTWGWVABIREPNRGTRLWL
 GTFNTSVEAAMAYDEAAKLYGHEAKLNLVHPQQQQQVNVNRNLSFSGHSGSGSWAYNKKL
 DMVHGLDLGLGQASCSRGSCSERSSFLQEDDDHSHNRCSSSSGSNLCWLLPKQSDSQDQE
 TVNATTSYGGEGGGSTLTFSTNLKPKNLMSONYGLYNGAWSRFLVGQEKTEHDVSSSC

GSSDNKESMLVPS CGGERMHRPELEERTGYLEMDDLLEIDDLGLLIGKNGDFKNWCCEEF
QHPWNWF*

>G2589 (23..1354)

AAAGAAAAGAAAAATAAAGATAATGAGGACGAAGACTAAGTTAGTACTCATACCTGATAG
ACACTTTCGGAGAGCCACATTCAGGAAGAGGAATGCAGGGATAAGGAAGAACTCCACGA
GCTGACAACTCTCTGTGACATCAAAGCATGTGCGGTAATCTACAGTCCGTTTCGAGAATCC
AACGGTGTGGCCGTCACCCGAAGGTGTTCAAGAGGTGATTTTCGGAGTTTCATGGAGAAGCC
GGCGACAGAACGGTCCAAGACGATGATGAGTCATGAGACTTTCTTGCGGGACCAAATCAC
CAAAGAACAAAACAACTAGAGAGTCTACGTCGTGAAAACCGAGAACTCAGCTTAAGCA
TTTTATGTTTGATTGCGTTGGAGGCAAGATGAGTGAGCAACAGTATGGTGCAAGGGACCT
TCAAGATTTAAGTCTTTTTACTGATCAATATCTTAATCAGCTTAATGCCAGGAAGAAGTT
CCTTACAGAATATGGTGAGTCTTCTTCTTCTGTTCTCTCTGTTTGATGTTGCGGGTGC
CAATCTCTCTGTTGTTGTCAGATCAAGCTGCGGTAAGTGTCTCTCTTGTGTTGCTGTTGC
GGGTGCCAATCTTCTGTTGTTGCTGATCAAGCTGCGGTAAGTGTCTCTCTGTTGCTGTTGC
TGTTGCGGGTGCCAACTCTCTGTTGTTGCTGATCAAGCTGCGGTAAGTGTCTCTCTGTTGCTGTTGC
ATTTTCATAACATGAATGTGAACCAGAATCAGTATGAGCCGGTTCAGCCCTATGTCCCTAC
TGTTTTTAGTGATCATATTCAATATCAGAATATGAACCTCAATCAAAACCAACAAGAGCC
GGTTCATTACCAGGCTCTTGCTGTTGCGGGTGCCGGTCTTCTATGACTCAGAATCAGTA
TGAGCCCGTTCACTACCAGAGTCTTGCTGTCGCGGGTGCGGCTTCTCTATGAGTCAGTT
GCAGTATGAGCCGGTTCAGCCTTATATCCCTACTGTTTTTAGTGATAATGTTCAATATCA
GCATATGAATTTGTATCAAAATCAACAAGAGCCGGTTCCTACCAAGCTCTTGGTGTTGC
AGGTGCCGGTCTTCTATGAATCAGAATCAGTATGAGCCGGTTCAGCCCTATGTCCCTAC
TGGTTTTAGTGATCATTTTTAGTTTGAGAAATATGAATTTGAATCAAAATCAACAGGAGCC
GGTTCAATACCAAGCTCTCTGTTGATTTTAATCATCAGATTCAACAAGGAACTATGATAT
GAATTTGAACCAGAATATGAGTTTGATGCCAAATCAGTATCCGTTTCAAAATGATCCATT
CATGAATATGTTGACAGAATATCTTATGAATAAGCGGGTATGTTGGAGAGCATGCAC

>G2589 Amino Acid Sequence (domain in AA coordinates: TBD)

MRTKTKLVLPDRHFRRATFRKRNAGIRKKLHELTTLCDIKACAVIYSPFENPTVWPSTE
GVQEVISEFMEKPAERSKTMMSHETFLRDQITKEQNKLESLRRENRETQLKHFMFDCVG
GKMSEQQYGARDLQDLSLFTDQYLNQLNARKKFLTEYGESSSSVPLFDVAGANPPVVAD
QAAVTVPPLFAVAGANLPVVADQAAVTVPPLFAVAGANLPVVADQAAVNVPFGFHNMNVN
QNQYEPVQPYVPTGFSDDHIQYQNMNFNQNQEPVHYQALAVAGAGLPMTQNQYEPVHYQS
LAVAGGGLPMSQLQYEPVQPYIPTVFSNVQYQHMNLQYQEPVHYQALGVAGAGLPMN
QNQYEPVQPYVPTGFSDDHFQFENMNLNQNQEPVQYQAPVDFNHQIQQGNYDMNLNQNM
LDPNQYPFQNDPFMNLTEYPYE*

>G2687 (45..1139)

CTCTGTCTCTCGTATCTTTCTACTACTCTGTTTCTTGAATTCTAATGAACAACATCGACG
ACGCAAGACGGAGACTTCAGTGTCTTCAGGTTCAAGCGACTCTTTCTTGCTCTCAAGA
AACGCATGAGACTTGATGACGAACCAGAAACGCCCTAGTGGTTTCGTCTTCACCAAGA
CGGTTGTGGCTTCTGGCAATGTCAAGTACAAAGGAGTCGTTAGCAACAGAACGGTCATT
GGGGTGCCAGATTTACGCAGACCACAAAAGGATTTGGCTTGAACCTTTCAAATCCGCTG
ATGAAGCCGCCACGGCTTACGATAGTGCATCTATCAAACCTCCGAAGCTTTGACGCTAACT
CGCACCGGAAC'TCCCTTGGTCTACAATCACTCTCAACGAACCAGACTTTCAAAATTGCT
ACACAACAGAGACTGTGTTGAACATGATCAGAGACGGTTCGTACCAACACAAATTCAGAG
ATTTTCTCAGAATCAGATCTCAGATTGTTGCGAGTATCAACATCGGGGGACCAAAACAAG
CCCGAGGAGAAGTGAATCAAGAATCAGACAAGTGTTTTTCTTGACACAGCTTTTTCAGA
AGGAATTGACACCGAGCGATGTAGGGAACTAAATAGGCTTGTGATACCTAAAAAGTATG
CAGTGAAGTATATGECTTTTCATAAGCGCTGATCAAAGCGAGAAAGAAGAGGGTGAAATAG
TAGGATCTGTGGAAGATGTGGAGGTTGTGTTTTACGACAGAGCAATGAGACAATGGAAGT
TTAGGTATTGTTACTGGAAAAGTAGCCAGAGCTTGTCTTACCAGAGGATGGAATAGTT
TCGTGAAGGAGAAGAATCTCAAGGAGAAGGATGTTATTGCCTTCTACACTTGCGATGTCC
CGAACATGTGAAGACATTAGAAGGTCAAAGAAAGAACTTCTTGATGATCGATGTTTCATT
GCTTTTCAGACAACGGTTCCTGTTAGCTGAGGAAGTAAGTATGACGGTTCATGACAGTT
CAGTGCAAGTAAAGAAAACAGAAAACCTTGGTTAGCTCCATGTTAGAAGATAAAGAAACCA
AATCAGAGGAGAACAAGGAGGGTTTATGCTGTTTGGTGTAAGGATCGAATGTCCTTAGG
GAATTTTTCTTTAAAAGTTTCTTACTTCAACTAGAACTTGTTTTACTTGTACCT

>G2687 Amino Acid Sequence (domain in AA coordinates: TBD)

MNNIDDAKTETSVSSGSSDSFLPLKKRMLDDEPENALVVSSSPKTVVASGNVYKGVVQ
QQNGHWGAQIYADHKRIWLGTFKSADEAATAYDSASIKLRSFDANSHRNFPWSTITLNEP
DFQNCYTTETVLNMIRDGSYQHKFRDFLRIRSQIVASINIGGPKQARGEVNQESDKCFSC
TQLFQKELTPSDVGKLNRLVIPKKYAVKYMPPFISADQSEKEEGEIVGSVEDVEVVFYDRA
MRQWKFRYCYWKSSQS FVFTRGWNSFVKEKNLKEKDVI AFYTC DVPNNVKTLEGQRKNFL
MIDVHCFSDNGSVVAEEVSM TVHDSSVQVKKTENLVSSMLEDKETKSEENKGGFMLFGVR
IECP*

>G27 (83..622)

CAAAATACCAAAAACAAAACATTTTTTTTAATCTTCCCACCAATTTTTTTCTCTTTCTCT
CGTTACATTAAATTATCTTTAGATGCAAGACTCTTCTCTCACGAATCGCAACGTAACCT
CCGGTCAACCGGTGCCGGAGAAAACCGGAAAGAGTTCTAAGACTAAAAATGAGCAAAAAGG
TGTTTCTAAACAACCAAAATTTTCGTGGGGTCAGAATGAGACAATGGGGAAAATGGGTGTC
TGAAATTAGAGAACCAAGAAAGAAATCAAGAATATGGCTCGGTACTTTCTCTACGCCGGA
GATGGCGGCGCGTGCACACGACGTGGCGGCTTTAGCCATCAAAGGTGGCTCTGCCCACCT
TAATTTCCCGGAGCTAGCTTACCATTGCGGAGACCGGCTAGCGCGGACCCTAAAGACAT
TCAAGAAGCCCGCCGCGCAGCAGCTGCCGTTGACTGGAAAGCACCGGAGTCTCCGTCTAG
CACCGTGACGTCTCTCCAGTCGCGGACGACGCTTCTCCGATCTTCTGATCTTTTGCT
TGACGTGAATGATCACAACAAAAACGATGGATTCTGGGACTCGTTTCCGTACGAAGATCC
TTTCTTCTTGAAAATTACTAGAAGGCAAATCTTGCCGGCGAACGGATTTCGGGTGGT
TTCCCGGTAAATAAGAAGACGATGTGCTTTTGACCTTTTGTCTACGATGGGAAATTT
CTTTTCTTTTACGTGTGAGTAAAGTTTCCGAATGTGTGATGTGTAAGTAAGTACAGGT
TATTTAATTTCTTTTTTTTGTACAAATACGTACGTATTACCAAAAAGTTTTCATTTATT
GTGCTTTTATCTTCCAAATTCATTAAAAAAAAAAAAAAAAAA

>G27 Amino Acid Sequence (domain in AA coordinates: 37-104)

MQDSSSHESQRNLRSPVPEKTGKSSKTKNEQKGVSKQPNFRGVRMRQWGWVSEIREPRK
KSRILWLTGFSTPEMAARAHDAALAIKGGSAHLNFPDELAYHLPRPASADPKDIEAAAAA
AAVDWKAPESPSSTVTSSPVADDAFSDLPLDLLLDVNDHNKNDGFWDSPFYEDPFFLENY*

>G2720 (1..894)

ATGGAAGCGAAGAAGAGATAAAGAAAGGTCCATGGAAAGCCGAAGAAGACGAAGTA
CTCATCAACCATGTCAAGAGATACGGTCTCTCGTGATTGGAGCTCCATTCGATCCAAAGGT
CTTCTTCAACGCACCGGCAAATCCTGTCTCTCGTTGGGTCAATAAACTCCGTCCCAAT
CTCAAAAATGGATGCAAGTTCTCGGCTGACGAAGAGAGGACTGTGATTGAGTTACAATCT
GAGTTTGGAACAAATGGGCGAGAATCGCTACGTATCTACCGGGAAGAACTGATAACGAT
GTGAAGAATTTCTGGAGTAGCAGACAAAAGAGACTCGCTAGGATTCTTCATAACTCCTCT
GATGCATCGAGTTCGAGTTTCAATCCCAAATCTTCTTCTCTCATCGACTCAAGGGCAAA
AACGTCAAACCAATCCGTCAATCCTCTCAGGGTTTTGGTTTTGGTTGAGGAAGAGGTTACA
GTTTCTTCTTCAATGTTCCAGATGGTTCCTTATTCATCTGATCAAGTTGGTGATGAAGTC
TTGAGGTTGCCGGATTGCGGTGTTAAGTTAGAGCATCAGCCTTTTCGCTTTTGGCACTGAT
CTTGCTCTAGCAGAGTACTCTGACTCACAGAATGATGCAATCAGCAAGCAATCAGCCCT
TTCTCTCCAGAAAGCAGAGAGCTTTTGCTAGACTTGACGACCCTTTTACTATGATATA
CTTGAGACAGCTGATTCTTCTGAGCCATTGTTGCTCTCCCTCAGCCGTTCTTCGAGCCT
TCGCCTGTGCCGAGAAGATGCAGACATGTTTCAAAGGATGAAGAAGCTGATGTTTCTTA
GACGATTTCCAGCTGACATGTTTGATCAGGTTGATCCAATCCCAAGTCCTTAG

>G2720 Amino Acid Sequence (domain in AA coordinates: 10-114)

MEAKKEEIKKGPWKAEEDEVLINHVKRYGPRDWSSIRSKGLLQRTGKSCRLRWVNLKLRPN
LKNGCKFSADEERTVIELQSEFGNKWARIATYLPGRTDNDVKNFWSRQKRLARILHNS
DASSSSFNPKSSSSSHRLKGNVKPIRQSSQGFGLVEEVTVSSSCSQMVYPYSSDQVGDEV
LRLPDLGVKLEHQPFAGTDLVLAEYSDSQNDANQQAISPFSPEARELLARLDDPFYDI
LGPADSSEPLFALPQFFEPSPVPRRCRHVSKDEEADVFLDDFPADMFDQVDPIPS*

>G2787 (142..1584)

TCTCAGAGCAAAAAACAAAAAAGAAAAAACCCTAAATCTAAATCTCACCTTCCA
CCTCTGTCTTTTTTTTTTTTGTCTTTTTTTTTTTTTTACTGTATCTTCTCTCTTTG
CTCTGCAAAAATCTCACATCCATGGATCCATCTCTTGGTGATCCTCATCATCTCCTCAG
TTCACCCCTTTTCTCATTTTCCACCTCCAATCATCATCTTTAGGACCAATCCGTAC
AATAACCATGTCCTCTTCCAACCGCAGCCGCAACGCAACGCAATCCCGCAACCGCAG
ATGTTTTCAGTTATCTCCACATGTTTCAATGCCCCACCCTCTTACTCCGAAATGATTG
GCTGCGATTGCGGCGTTAAACGAACCGGATGGTTCGAGCAAGATGGCAATTCGAGATAC

ATCGAGAGATGTTACACCGGTTTAACTTCTGCTCATGCTGCTTTGTTGACTCACCATCTC
 AAGACTTTGAAGACCAGTGGTGTCTTTCTATGGTTAAGAAATCTTACAAAATTGCTGGT
 TCTTCTACTCCTCCTGCTAGTGTAGCTGTTGCTGCTGCTGCCGCCGCTCAAGGTCTCGAT
 GTTCCCAGATCTGAGATTCTCCATTCAAGTAACAACGATCCCATGGCTTCTGGCTCTGCT
 TCTCAGCCTCTGAAACGAGGTGCTGGTGTCTCTCTAAGCCTAAACCTGAATCTCAACCA
 CAACCACTACAGCAACTTCCACCGACCAATCAAGTCCAGGCTAACGGACAGCCAATCTGG
 GAACAGCAGCAAGTTCAATCACCTGTTCGGTTCCGACTCCGGTTACAGAGTCGGCGAAG
 AGAGGACCTGGTGTCTCAAGGAAGAACGGTTCTGCTGCTCTGCTACTGCACCAATCGTT
 CAAGCTTCGGTTATGGCTGGAATTATGAAACGTAGAGGTAGACCACCGGGTGTGAGCT
 GCTGGGAGACAGAGGAAGCCCAAATCCGTTTCTTCTACTGCCTCTGTGTATCCTTATGTT
 GCTAATGGTGTAGACGCAGAGGAAGGCCTAGGAGAGTTGTTGACCCTAGCAGTATTGTT
 AGTGTGTGCTCCAGTAGGTGGTGAATGTGGCAGCGGTTGCCGCCAGGGATGAAGCGTGGA
 CGTGGACGACCACCTAAGATTGGTGGTGTATCAGTAGGCTTATTATGAAGCCTAAGAGA
 GGACGAGGACGCTCTGTAGGTAGACCCAGAAAGATTGGAACATCAGTCACGACTGGGACA
 CAAGATTCTGGAGAACTCAAGAAGAAGTTTGATATTTTCAAGAGAAAGTGAAGAAATT
 GTGAAGGTGTTGAAGGATGGAGTTACAAGTGAGAATCAAGCAGTGGTGCAAGCCATAAAA
 GATCTGGAAGCACTAACAGTGACGGAGACCGTTGAGCCACAAGTTATGGAAGAAGTGCAG
 CCAGAGGAGACTGCAGCACCACAGACTGAAGCTCAACAACTGAAGCTGCTGAGACACAA
 GGAGGACAAGAAGAAGGACAAGAAAGAGAAGGAGAAACACAGACCCAGACAGAAGCAGAG
 GCAATGCAAGAAGCTCTGTTCTGAAGAATAATAATGATCTAGAAAACAACCTAGACATAA
 TAGCCTTGGTGTGTTGGCGTTAGGAGTGTGTTTTTTTTTAGTTGTTTAGGTGTTGGAATCGC
 ATCTTAAATTATATAAAAATCTATAAGGAATTTTAAATTTTCTAGGTTTTGTTGTCTGCA
 GAAGAAGAAATAGTAGACTCGTTAATGGTGTGTTGTGCGGTGTGCTTTAACCAAACCAT
 AAGACGTGGCTGTAAATTACGATGTTTCTAGTCTTCCATCTTTAATAATCTCTTATTGC
 GTCTGTGCTTTGTTTTT

>G2787 Amino Acid Sequence (domain in AA coordinates: 172-192, 226-247, 256-276, 290-311, 245-366)

MDPSLGDPHHPQFTFPFPHFTSNHHPLGPNPYNHVVFPQPQTQTQIPQPQMFQLSPH
 VSMPPHPYSEMICAIAALNEPDGSSKMAISRYIERCYTGLTSAHAALLTHHLKTLKTSG
 VLSMVKKSYKIIAGSSTPPASVAVAAAAAQGLDVPRSEILHSSNNDPMASGSASQPLKRG
 RGRPPKPKPESQPQPLQLPPTNQVQANGQPIWEQQQVQSPVPVPTPVTESAKRGPGRPR
 KNGSAAPATAPIVQASVMAGIMKRRGRPPGRRAGRQRKPKSVSTASVYPYVANGARRR
 GRPRRVDPSSIVSVPVGGENVAAPGPKRGRGRPPKIGGVISRLIMPKRGRGRPVG
 RPRKIGTSVTTGTQDSGELKKKFDIFQEKVKEIVKVLKDGVTSENQAVVQAIKDLEALTV
 TETVEPQVMEEVQPEETAAPQTEAQTEAAETQGGQEEGQEREGETQTQTEAEAMQEALF
 *

>G2789 (82..879)

CTTTAGGGACACCAATCTATTCAACCTAAAAGCCTTCTTTTCCCTATATTGACCAACT
 TTTTAGCGAATCAGAAGAGGAATGGATGAGGTATCTCGTTCTCATACACCGCAATTTCTA
 TCAAGTGATCATCAGCACTATCACCATCAAAACGCTGGACGACAAAAACGGGCAGAGAA
 GAAGAAGGAGTTGAACCAACAATATAGGGGAAGACCTAGCCACCTTTCCTTCCGGAGAA
 GAGAATATCAAGAAGAGAAGGCCACGTGGCAGACCTGCTGGTTCCAAGAACAACCCAAA
 GCACCAATCATAGTCACTCGCGACTCCGCGAACGCCTTCAGATGTCACGTATGGAGATA
 ACCAACGCCTGCGATGTAATGGAAAGCCTAGCCGTCTTCGCTAGACGCCGTGAGCGTGGC
 GTTTGCGTCTTGACCGGAAACGGGGCCGTTACAAACGTACCGTTAGACAACCTGGCGGA
 GGCGTCGTGATTTACACGGACGGTTTGAGATTCTTTCTCTCTCGGGTTCGTTTCTTCCT
 CCACCGGCACCACAGCTGCGTCTGGTTTAAAGGTTTACTTAGCCGGTGGTCAAGGTCAA
 GTGATCGGAGGCAGTGGTGGGACCGCTTACGGCATCAAGTCCGGTGGTGTGTTATGGCA
 GCTTCATTTGAAACGCATCTTACGAGAGGCTGCCACTAGAGGAGGAGGAGGAACTGAA
 AGAGAAATAGATGGAAACGCGGCTAGGGCGATTGGAACGCAACGCAGAAACAGTTAATG
 CAAGATGCGACATCGTTTATTGGGTGCGCGTCGAATTTAATTAACCTCTGTTTCGTTGCCA
 GGTGAAGCTTATTGGGGAACGCAACGACCGTCTTTCTAAGATAATATCATTGATAATATA
 AGTTTCGTCTTCTATTCTTTTCACTTTTACCTTTTCACTTTCTTAGGTTTTGTTTT
 AACGTTTGATTAATACCTGAAGGTTTTTGAAAATTTTCGATCGGATAAAGGATTTATG
 TTGCGAGCCGAAACGCGGCC

>G2789 Amino Acid Sequence (domain in AA coordinates: 53-73, 121-165)

MDEVSRSHTPQFLSSDHQHYHHQNAQRQKRGREEGVPEPNIGEDLATFPSGEENIKRR

PRGRPAGSKNPKAPIIVTRDSANAFRCHVMEITNACDVMEISLAVFARRRQRGVCVLTGN
GAVTNVTVRQPGGGVVSLHGRFEILSLSGSFLPPPAPPAASGLKVYLAGGQGVIGGSV
GPLTASSPVVMAASFGNASYERLPLEEEEEETEREIDGNAARAIGTQTQKQLMQDATSFI
GSPSNLINSVSLPGEAYWGTQRPSF*

>G31 (13..615)

CTTTTATAAGCAATGGCTCCAAGACAGGCGAACGGTAGAAGCATTGCCGTGAGTGAAGGC
GGCGGAGGGAAGACGATGACGATGACGACGATGCGGAAGGAAGTGCACCTTAGAGGTGTG
AGGAAGCGTCCATGGGGTAGATACGCGGCGGAGATCCGTGACCCGGGAAAGAAAACCCGG
GTTTGGCTCGGGACATTTCGACACGGCGGAGGAAGCTGCAAGAGCTTACGACACCGCCGCT
AGAGAGTTTCGTGGCTCCAAAGCAAAGACTAATTTCCCTCTTCCCGGAGAGTCTACTACG
GTTAACGACGGTGGCGAGAACGATTCTTACGTCAACCGTACGACGGTGACGACGGCGCGT
GAGATGACCGCTCAGAGATTTCCGTTTGCATGTCAACGGGAGCGTAAAGTCGTGCGTGGT
TATGCTTCTGCTGGTTTTTTCTTCGATCCCGTCAAGAGCTGCTTCGTTAAGAGCAGAGCTT
TCTCGGGTTTGTCCGGTTCGGTTTGATCCGGTTAATATCGAGTTGAGTATTGGTATTCTGA
GAAACCGTAAAAGTTGAACCGAGAAGAGAACTAAACCTGGATCTTAACCTAGCTCCACCG
GTGGTGGACGTTTAGATTTTTTTCTTTTTCATAATTTGTATTTTACATTGCCGAAAA
TAATTAATGTTTTCTTTAG

>G31 Amino Acid Sequence (domain in AA coordinates: TBD)

MAPRQANGRSIAVSEGGGKMTMTTMRKEVHFRGVRKRPWGRYAAEIRDPGKKTRVWL
TFDTAEAAARAYDTAAREFRGSKAKTNFPLPGESTTVNDGGENDSYVNRRTVTAREMTR
QRFPPACHRRERKVVGGYASAGFFDPSRAASLRAELSRVCPVRFDPVNIELSIGIRETVK
VEPRRELNLNLAPPVVDV*

>G33 (20..757)

ATTCTCCCCAACCAAAATATGACCACAGAAAAAGAGAATGTCACTACGGCCGTGGCCGT
GAAAGACGGCGGAGAAAAGAGTAAGGAAGTGAGTGACAAGGGCGTAAAGAAGAGAAAGAA
TGTAACATAAGGCCCTGGCCGTGAATGACGGCGGAGAAAAGAGTAAGGAAGTGCGTTACAG
GGGTGTAAGGAGGAGACCATGGGGGAGATATGCTGCGGAGATCCGTGATCCGGTAAAGAA
AAAACGGGTCTGGCTCGGGTCTTCAACACGGGGGAGGAAGCCGCCAGAGCCTACGACTC
CGCTGCCATAAGGTTTCGAGGATCGAAAGCTACTACTAACTTCCCTCTAATCGGATACTA
TGGGATTTCTTCGGCGACGCCGGTGAACAACCAACCTTCCGAGACGGTGAGTGATGAAAA
TGCCAACTCCCTCTCGTTGGAGACGATGGGAATGCTTTGGCTTCTCCGGTGAACAACAC
CTTTCGGAACGGCGCGTGTATGGAACACTTCCATCGGATTGTACGACATGTTATCTCC
GGGGTGGCTGAAGCGGTGTCTGGATTTTTCTTAGATCTGCCTGAAGTTATTGCGTTGAA
AGAGGAGCTTGATCGAGTTTGTCTGACAGTTTGAGTCCATTGATATGGGGTTGACTAT
TGGTCTCTCAAACCGCGTGGAAGAGCTGAGACTTCCCTCCGCCGTGGATTGTAAGCTGCG
AATGGAACCGGATCTTGACCTCAACGCAAGTCCCTAAAGATTGATCTGATGTTGTTGTCC
TTGAATAAGTTTGTATCTTGTCTCTTCTGATTGTCTGTACTTCTATTGGTTGATTGCG
TGCTTTTGGAGGACAAAAACAACATTTTTTTATGTATTAATAAAGGTAATTGAACTATT
ATCGAAAAA

>G33 Amino Acid Sequence (domain in AA coordinates: 50-117)

MTTEKENVTAVAVKDGGEKSKEVSDKGVKRKNVTALAVNDGGEKSKEVRYRGVRRRP
WGRYAAEIRDPVKKRVWLGSFNTGEEAARAYDSAAIRFRGSKATTNFPPLIGYYGISSAT
PVNNNLSETVSDGNANLPLVGDDGNALASPVNNTLSETARDGTLPSDCHDMLSPGVAEAV
AGFFLDLPEVIALKEELDRVCPDQFESIDMGLTIGPQTAVEEPETSSAVDCKLRMEPDLD
LNASP*

>G342 (1..723)

ATGGACGTCTACGGCATGTCTTACCAGACTTGCTTCGTATCGACGACCTTCTCGATTTC
TCCAACGACGAAATCTTCTCTTCTCTTCCACCGTCACTTCCCTCCGCCGCTTCTCCGCC
GCTTCTTCCGAAAACCTTTTACGCTTCTTCTTCCACCTACACTTCTCCTACTCTCCTC
ACCGACTTCACTACGATCTCTGCGTTCAGTACGACGACGCTCATCTCGAATGGTTA
TCGCGATTCTGTTGACGATTCACTTCCGATTTCAGCAATCCTTTAACCATGACCGTT
AGACCCGAGATTTCACTACCGGAAAACCTAGAAAGTCGCGGATCAAGAGCACCAGCACCT
TCCGTAGCTGGAATCTGGGCTCCGATGTCTGAATCAGAGCTTTGTCACTCCGTCGCTAAA
CCTAAACCGAAGAAAGTCTACAACGCTGAATCGGTTACGGCGGATGGAGCGAGGCGGTGC
ACGCACTGTGCTCGGAGAAAACGCCACAGTGGAGAACTGGACCGCTTGACCTAAAACA
CTTTGTAAACGCTTGTGGAGTTCTGTACAAATCAGGGAGGCTTGTACCGGAATACAGACCG
GCGTCGAGTCCGACGTTTGTATTGACTCAGCATTCGAACTCTCATCGGAAAGTTATGGAG

>G342 Amino Acid Sequence (domain in AA coordinates: 155-190)
MDVYGMSSPDLLRIDDLLDFSNDEIFSSSSTVTSSAASSAASSENPFSPSPSTYTSPTLL
TDFTHDLCVPSDAAHLEWLSRFVDDSFSDFPANPLTMTVRPEISFTGKPRRRSRAPAP
SVAGTWAPMSESELCHSVAKPKPKKYNAESVTADGARRCTHCASEKTPQWRTGPLGPKT
LCNACGVRYKSGRLVPEYRPASSPTFVLTHQHSNSHRKVMELRRQKEQQESCVRIPIPFQPO
*

>G352 Amino Acid Sequence (domain in AA coordinates: 99-119,166-186)
MALETLSNPTATTTARPLRLRYREEMEPENLEQWAKRKRTKRQRFDHGHQNETNKNLPSE
EEYLALCLLMLARGSAVQSPPLPLPSRASPSDHRDYKCTVCGKSFSSYQALGGHKTSR
KPTNTSITSGNQELSNNSHSNSGVSVINVTVNTGNGVVSQSGKIHTCSICFKSFASGQALG
GHKRCHYDGGNGNGNGSSNSVELVAGSDVSDVDNERWSEESAIGGHRGFDLNLNLPADQV
SVTTS*

ATGTCAGAACAACACAAATGCAAGCTCTGTTCCAAGAGTTTCTGTAATGGCAGAGCACTT
GGTGGTCACATGAAGTCTCACTTGGTCTCATCTCAGTCTTCAGCTCGGAAGAACTAGGT
GACTCGGTCTATTCTTCTTCTTCTTCTTCCGATGGTAAAGCGCTCGCCTACGGGTTA
CGAGAGAACCCGAGGAAGAGTTTCCGGGTCTTTAATCCGGATCCTGAGTCATCCACAATT
TACAACAGTGAGACAGAGACCGAACCTGAATCCGGAGACCCGGTTAAGAAACGGGTCAGA
GGAGATGTTTCAAAGAAGAAGAAGAAGGCAAAGAGTAAGAGAGTGTTTGAGAACTCG
AAGAAGCAAAAGACAATTACAGAGTCACCAGAACCAGCGAGTTCTGTCTCTGATGGTTCT
CCTGAACAAGATTTAGCTATGTGCTTGATGATGCTGTCAAGAGATTCAAGGGAGCTCGAG
ATTAACCTGAAAAAACCGGAGGAAGAGAGGAAGCCGGAAAAAGACATTTCCCTGAGCTC
CGTTCGTGTATGATAGATCTGAATCTTCTCCGCCGCAAGAAGCTGAAGAGCTGTACCCGTC
GTTTCAGCCATATAA

26397 Amino Acid Sequence (domain 1) (11-0001) (11-0001)
 MQNKHCKLCSKSF CNGRALGGHMKSHLVSSQSSARKKLGDSVSSSSSSSDGKALAYGL
 RENPRKSFrvfnpdpessttynsetetepesgdPVKKVRGdVSKKKKKAKSKRVFENS
 KKQKTIHESPEPASSVSGSPeqDLAMCLMMLSRDSRELEIKLKKPEEBERKPEKRHFPEL
 RRCMIDLNLPPPOEAeAVTVVSAI*

ATGGGTCAAGATGAGGTTGGGAGTGATCAGACGCAAATCATAAAAGGGAAACGTACGAAG
CGACAAAGATCGTCTTCGACGTTTGTGGTGACGGCGGCGACAACAGTGACTTCAACAAGT
TCATCGGCCGGTGGAAGTGGAGGAGAAAGAGCTGTTTCAGATGAATACAACCTCGGCGGTT
TCGTCTCCGGTGACTACTGATTGTACGCAAGAAGAAGACATGGCGATTTGTCTCATC
ATGTTAGCTCGTGGGACAGTTCCTCCATCGCCGGATCTCAAGAACTCGAGAAAAATTTCAT
CAGAAGATTTTCGTCCGAGAATTCTAGTTTCTATGTGTACGAGTGTA AACCGTGTAACCGG
ACGTTTTTCGTGCTTCCAAGCACTTGGTGGACACAGAGCGAGCCACAAGAAGCCGAGGACG
TCGACTGAGGAAAAGACTAGACTACCCCTGACGCAACCAAGTCTAGTGACATCAGAAGAA
GGGCAAAAACAGTCATTTCAAAGTTTCCGGCTCAGCCCTAGCTTCA CAGGCAAGTAACATC
ATCAAC AAGGCAAAACAAAGTACACGAGTGTTCCATCTGCGGTTCTGAGTTCACTTCCGGG

CAAGCTCTCGGTGGTCACATGAGGCGGCACAGGACAGCCGTAACCACGATTAGCCCCGTT
GCAGCCACCGCAGAAGTAAGCAGAAACAGTACAGAGGAAGAGATTGAGATCAATATAGGC
CGTTCGATGGAACAGCAGAGGAAATATCTACCGTTGGATCTTAATCTACCAGCACCAGAA
GATGATCTAAGAGAGTCAAAGTTTCAAGGGATAGTATTCTCAGCAACACCAGCGTTAATA
GATTGTCATTACTAG

>G358 Amino Acid Sequence (domain in AA coordinates: 124-135, 188-210)

MGQDEVGSDQTOI IKGKRTKRQRSSSTFVVTAATTVTSTSSSAGGSGGERAVSDEYNSAV
SSPVTTDCTQEEEDMAICLIMLARGTVLPSPDLKNSRKIHQKISSENSSFYVYECKTCNR
TFSSFQALGGHRASHKKPRTSTEEKTRLPLTQPKSSASEEGQNSHFVSGSALASQASNI
INKANKVHECSICGSEFTSGQALGGHMRHRHTAVTTISPVAATAEVSERNSTEEIEINIG
RSMQQQRKYLPLDLNLPAPEDDLRESKFQGIVFSATPALIDCHY*

>G360 (1..543)

ATGTGGAACCCTAACAAAATTGAAGAATTGGAGGATGATGATGAATCTTGGGAAGTCAAA
GCCTTTGAGCAAGACACTAAAGGCAACATCTCTGGTACCACTTGGCCTCCAAGATCTTAC
ACTTGCAATTCTGCCGCCGTGAGTTCGGTTCTGCTCAAGCCTTAGGCGGTACATGAAT
GTCCACCGCGGTGACCGCGCCTCATCTAGGGCTCATCAAGGTTCCACCGTTGCGGCTGCG
GCTAGAAGCGGCCACGGGGGGATGTTACTCAATTCTGTGCTCCGCCGTTGCCTACAACG
ACACTTATAATACAATCCACGGCGAGTAACATTGAAGGTTTGTCCCATTCTACCAACTG
CAAAACCCTAGTGGCATTCTTGGTAATTCTGGTGACATGGTGAATCTTTATGTAGAAGTT
CCTCCTCGGCTTATTGAATATTCGACAGGAGATGATGAGAGCATTGGCTCGATGAAAGAA
GCGACAGGAACATCAGTGGATGAGCTTGATCTTGAACCTTCGGCTAGGGCACCATCCACCG
TGA

>G360 Amino Acid Sequence (domain in aa coordinates: 42-62)

MWNPNKIELEDDDESWEVKAFQDTKGNISGTTWPPRSYTCNFCRREFRSAQALGGHNM
VHRRDRASSRAHQGSTVAAAARSGHGMLLNSCAPPLPTTLIIQSTASNIEGLSHFYQL
QNPSGIFGNSGDMVNLYVEVPPRLIEYSTGDDESIGSMKEATGTSVDELDELRLGHHP
*

>G362 (195..830)

ATAAAAAACCTTCATACAATATAAAATTTCTTTAGACATACAATATATTATACTATTAC
AGATGCAATGCATCATTAGTTACAAACTATTAACTAAATATCCCCGCTCTCTCTTGC
TATATAAAGAAGATCATTTACACATCTCCTTAAGCAAATTAAACCCATCGATAAACACAT
ACGTTTCACACATATATGTCTATAAATCCGACAATGTCTCGTACTGGCGAAAGTTCTTCAG
GTTTCGTCTCCGACAAGACGATAAAGCTATTTCGGCTTCGAACTCATCAGCGGCAGTCGTA
CGCCGGAAATCAGACGGCGGAAAGCGTGAGCTCGTCCACAAACACGACGTCGTTAACAG
TGATGAAAAGACACGAGTGCCAATACTGCGGTAAAGAGTTTGCAAATTCTCAAGCCTTAG
GAGGTCACCAAAACGCTCACAAGAAGGAGAGGTTGAAGAAGAAGAGGCTTCAGCTTCAAG
CTCGGCGAGCCAGCATCGGCTATTATCTCACCAACCACCAACAACCCATAACGACGTCAT
TTCAGAGACAATACAAAACGCCGTCGTATTGTGCATTCTCCTCCATGCACGTGAATAATG
ATCAGATGGGTGTGTACAACGAAGATTGGTCGTGAGGTCGTGCGAGATTAACTTCGGTA
ATAATGACACGTGCCAAGATCTTAATGAACAAAGCGGTGAGATGGGTAAGCTGTACGGTG
TTCGACCGAACATGATTCACTTCCAGAGAGATCTGAGTTCTCGTTCTGATCAGATGAGAA
GTATTAACTCGCTGGATCTTCATCTAGGTTTTGCCGGAGATGCGGCATAACAAATTAAAG
AGAGATATATGATTAAAGATTATATGTACTATAGTGGCGTATTTTCATTGGGATCATGAAGG
GGAAAAACGAGACATATAGTATTCTTGATGCAATTTGAGTTTTGTAATTTATTTAGGTT
TATGTATGTTTTTCGAAG

>G362 Amino Acid Sequence (domain in AA coordinates: 62-82)

MSINPTMSRTGESSSSGSSSDKTIKLFGEFELISGSRTPETITAEVSSSTNTTSLTVMKRH
ECQYCGKEFANSQALGGHQNAHKKERLKKRLQLQARRASIGYYLTNHQQPITTSFORQY
KTPSYCAFSSMHVNNDQMGVYNEDWSSRSSQINFGNNDTCQDLNEQSGEMGKLYGVRPNM
IQFQRLSSRSQDQMRINSLLDLHLGFAGDAA*

>G364 (64..516)

AAGCTTGATATCGCCTCTCTCTAATCTCTCTTTCTCTCTCTATCTCTAAGAATATATAAA
GGTATGGACTACCGACCAACACACATCCCTACGTCTAAGCCTACCAAGTTACAAGAACCAC
CAACTAAACCTAGAAGTTGTTCTCGAGCCTTCTTCCATGTCTTCTTCTTCATCTTCTTCC
ACGAACATCATCATGTTTGGAGCAGCCTAGGGTATTCTCATGTAAGTATTGTCAAAGA
AAGTTTTACAGCTCTCAAGCTCTTGGTGGTCATCAAACGCTCATAAGCTTGAGAGAACC
TTAGCCAAGAAGAGTCGAGAACTCTTTAGATCCTCAAACACTGTTGATTCTGATCAGCCT

TACCCGTTCTCCGGTTCGCTTTGAGCTTTACGGCCGTGGCTACCAAGGATTTCTCGAAAGT
GGCGGCTCGAGGGACTTCTCCGCCCCGCGTGTGCCGGAGAGTGGTCTTGATCAGGATCAG
GAGAAGAGTACCTTGACTTATCCTTAAGGCTCTAAAAGAATCTTATATTTTGTAGTCT
ATATATTATCATATCAATTGTTAATCTTAAAATTGATTGTTTTACTTATTAGTCATTTCC
TATTATCTGAAAGTTTTCTTTGTAAGTTGTAAGTATGGTCCTAAATTCAAATCCAAATTT
GATTTTGAAGATGGTACCTAATGCAGTAGTTAAATAAGTTAAAAAATGAAGGATCTAT
AATTCTCT

>G364 Amino Acid Sequence (domain in AA coordinates: 54-76)
MDYQPNTSLRLSLPSYKNHQLNLELVLEPSSSSSSSSSTNSSSCLEQPRVFSCNYCQRK
FYSSQALGGHQNAHKLERTLAKKSRELFRSSNTVDSQPYPFSGRFELYGRGYQGFLSESG
GSRDFSARRVPESGLDQDQEKSHLDLSLRL*

>G365 (69..755)
CAATCTTTTACTTTTCATTCTCTTTATATATTCTCTCTACGCTATAATATATATTACACA
GAATATACATGGAACCGTCCATCAAAGGAGATCAAGAAATGTTAAAAATCAAGAAACAAG
GTCATCAAGATCTTGAGTTGGGGTTGACCTTTTGTACGTGGAACCGCGACCTCATCAG
AGCTCAATCTCATCGATTCTTTCAAACACAGCTCATCATCGACTTCTCATCATCAGCACC
AGCAAGAACAATTGGCAGATCCGAGAGTGTCTCGTGTAAATTATTGTCAAAGAAAGTTCT
ATAGTTTACAAGCGCTAGGCGGTACCAAAACGCTCATAACGTGAGCGCACCTTAGCCA
AACGTGGACAGTATTACAAGATGACTCTCTCCTCCTTGCCCTTCTTCAGCGTTTGCCTTTG
GCCACGGTTTCAGTCAGCAGATTTCGCAAGCATGGCATCGTTACCATTACATGGCTCGGTGA
ATAACAGGTCAACGTTAGGGATTCAAGCTCATTCAACGATCCATAAGCCCAGCTTCTTAG
GAAGACAAACGACGAGTTTAAAGTCATGTTTTCAAACAGAGCATTACCAGAAACCGACCA
TAGGAAAGATGTTGCCGGAGAAATTTACCTTGAAGTCGCCGGAATAATAACAGTAACA
TGGTTGCTGCTAAGTTGGAGAGAATTGGACATTTCAAGAGCAACCAAGAAGATCATAATC
AGTTTAAAGAAATTGACTTGACTCTTAAAGCTATGAGCTCTGCCATCTTCTTTTAGTCTT
CATTATAACTTTTTTTATTCTCATCTTTGTTTGATATAATGATTGACGGCAGGGTGTGTT
AGAGTTTCACTAATGATCAAGTTGTACTTTTTATATATTTCATTGATACCTTGTGTGATG
AATTCAATATTTTAGGTCTGTTTTT

>G365 Amino Acid Sequence (domain in aa coordinates: 70-90)
MEPSIKGDQEMLKIKKQGHQDLELGLTLLSRGTATSSSELNLIDSFKTSSSSSTSHHQHQQE
QLADPRVFSCNYCQRKFYSSQALGGHQNAHKRERTLAKRGQYYKMTLSSLPSAFAFGHG
SVSRFASMASLPLHGSVNNRSTLGIQAHSTIHKPSFLGRQTTLSLHVFKQSIHQKPTIGK
MLPEKFHLEVAGNNNSNMVAALKERIGHFKSNQEDHNQFKKIDLTLLKL*

>G367 (1..708)
ATGGACGCTTCAATAGTTTCTCATCCACTGCTTTTCCATATCAAGATTCTCTAAACCAG
AGCATCGAAGACGAGAAAGAGACGTTTATAATTCTAGTCACGAACCTCAATCTCATCGAC
TGCATAGACGACACAACGAGTATCGTTAACGAATCTACAACATCCACAGAACAAAAGCTT
TTCTCATGCAACTATTGTCAAAGAACTTTCTATAGCTCACAAGCACTTGGTGGTCACCAA
AACGCACACAAGAGAGAGAGAACGTTGGCGAAGAGAGGACAACGTATGGCAGCGTCAGCC
TCAGCTTTTGGACATCCTTACGGTTTCTCTCCACTTCTTTCCACGGACAGTACAACAAC
CATAGGTCTTTAGGGATCCAAGCGCATTCGATAAGCCACAAGCTAAGTTCTTATAACGGG
TTTGGTGGTCACTATGGTCAGATCAACTGGTCAAGACTTCCATTGTATCAACAACAGCC
ATAGGTAAATTTCCCTCAATGGATAATTTTCATCATCATCATCATCAGATGATGATGATG
GCTCCTTCAGTAAATTCACGGTCCAATAACATCGATAGCCCAAGCAACACAGGACGGGTT
CTAGAAGGGTCACCGACTCTTGAACAATGGCACGGAGACAAAGGATTGTTGTTAAGTACA
AGTCATCATGAAGAGCAGCAGAAACTTGACTTGTCCCTCAAGCTTTGA

>G367 Amino Acid Sequence (domain in AA coordinates: 63-84)
MDASIVSSSTAFPYQDSLNSIEDEERDVHNSSELNLIDCIDDTTISIVNESTTSTEQKL
FSCNYCQRTFYSSQALGGHQNAHKRERTLAKRGQMAASASAFGHYPYGFSPLPFHGQYNN
HRSGLGIQAHSSISHLSSYNGFGGHYQINWSRLPFDQQAIGKFPMDNFHHHHHQMmmm
APSVNSRSNNIDSPSNTGRVLEGSPTLEQWHGDKGLLLSTSHHEEQKLDLSLKL*

>G373 (1..1854)
ATGGCGATTGAAACTCAGCTTCTCTTGGCAGCGTGACGGTGTGTGTATGCGGTGTCAGGTG
AATCCTCCGTCGAAGAGACTCTCACTTGTGGCAGCTGCGTCACTCCATGGCACGTGCCG
TGTCTCCTCCCCGAATCACTCGCTTCTTCCACTGGAGAGTGGGAGTGTCCCGATTGCTCC
GGCGTTGTGCTTCCCTCCGCCGTCCGGGTACCGGAAACGCTCGACCTGAATCTTCCGGT
TCAGTTCTCGTTGCTGCGATCCGTGCGATTACGGCTGATGAGACTTTAACCGAAGCTGAG

AAAGCCAAAAAAGGCAGAACTGATGAGTGGGGGTGGTGACGATGGTGTGCGATGAAGAA
GAGAAGAAGAAGTTAGAAATCTTTTGTCTATTTGCATTCAATTGCCAGAAAGACCTATC
ACGACACCGTGTGGGCACAATTTCTGTTTGAAATGTTTCGAGAAATGGGCAGTAGGTCAA
GGGAAGCTAACTTGATGATATGCCGAAGCAAAATTCGAGACATGTGGCAAAAAATCCT
CGCATCAACTTAGCTCTAGTTTCTGCTATTCGTTTAGCAAATGTTACCAAATGTTCTGTT
GAGGCAACTGCAGCCAAGGTTTCATCATATTATCCGCAACCAAGACCGTCTGAGAAAGCA
TTTACTACCGAGCGGGCAGTAAAACTGGGAAAGCTAATGCTGCTAGCGGTAAGTTTTTT
GTGACAATACCTCGTGATCATTTTGGTCCCATACCAGCTGAGAATGATGTCACTAGAAAG
CAAGGTGTTTTGGTTGGAGAATCTTGGGAGGACAGGCAAGAGTGTAGGCAGTGGGGAGCT
CATTTCGCCCATATTGCTGGCATTGCCGGCAATCAGCGGTTGGAGCTCAGTCTGTGGCC
CTCTCTGGAGGTTATGACGATGATGAGGATCATGGTGAATGGTTTCTCTACACAGGAAGT
GGTGGAAAGGGATCTCAGTGGAAACAAAAGAATTAACAAGAAACAGTCGTCTGACCAGGCG
TTTAAAAACATGAATGAATCTCTAAGACTTAGTTGCAAAATGGGCTATCCTGTCCGAGTT
GTCAGGTCTTGAAGGAGAAGCGTTCTGCATATGCCCTGCTGAAGGTGTGAGATATGAT
GGGGTCTATCGAATTGAGAAGTGTGGAGTAATGTTGGAGTACAGGGTTCTTTAAGGTC
TGTCGTTACCTGTTTGTAGATGTGACAATGAGCCAGCTCCATGGACCAGTGATGAGCAT
GGCGATCGTCCAAGACCGTTGCCAATGTTCCGGAGCTTGAGACTGCTGCTGACCTGTTT
GTGAGAAAGGAGAGTCCATCATGGGATTTTCGATGAAGCTGAGGGTCTGTGGAAATGGATG
AAGTCTCCTCTGTTAGCAGAATGGCTTTGGATCCTGAGGAGAGGAAGAATAAGAGA
GCAAAAAATACTATGAAGGCCAGACTTCTGAAAGAATTTAGTTGCCAAATCTGTCCGGAA
GTGCTGAGTCTTCCAGTGACGACGCCTTGTGCACACAACCTCTGCAAGCATGCTTAGAA
GCGAAGTTTGTGGGATAACTCAACTGAGAGAGAGAAGCAATGGCGGACGTAAACTACGT
GCAAGAAGAACATCATGACCTGCCCTTGTGCACGACGGATCTCTCCGAGTTTCTCCAA
AACCCGCAGGTGAACAGAGAGATGATGGAGATAATAGAGAATTTAAGAAGAGTGAGGAA
GAGGCTGATGCATCCATTTCTGAAGAAGAAGAAGAATCCGAACCTCCAACATAAGAAG
ATTAAGATGGATAACAACCTCTGTTGGTGGTAGTGGTACAAGTCTCTCAGCTTAA
>G373 Amino Acid Sequence (domain in AA coordinates: 129-168)
MAIETQLPCDGDGVCMRCQVNPSEETLTCGTCVTPWHVPCLLPESLASSTGEWECPCDS
GVVVPSAAPTGNARPESGSLVLAIRAIQADETLTEAEKAKKRQKLMSGGDDGVDEE
EKKKLEIFCSICIQLPERPITTPCGHNFLCKCFEKWAVGQKLTMCICRSKIPRHVAKNP
RINLALVSAIRLANVTKCSVEATAAKVHHIIRNQDRPEKAF'TTERAVKTGKANAASGKFF
VTIPRDHFGPIPAENDVTRKQGVLVGESWEDRQECRQWGAHFPHIAGIAGQSAVGAQSV
LSGGYDDDEDHGEWFLYTGSGGRDLGSKNRINKKQSSDQAFKNMNESSLRLSCKMGYPVRV
VRSWKEKRSAYAPAEGVRYDGVYRIEKCWSNVGVQGSFKVCRYLFVRCNDNEPAPWTSDEH
GDRPRPLPNVPELETAADLFVRKESPSWDFDEAEGRWKWMKSPVSRMALDPEERKKNKR
AKNTMKARLLKEFSCQICREVLSPVTTCAHNFCACLEAKFAGITQLRERSNGGRKLR
AKKNIMTCCCTTDLSEFLQNPQVNREMEIENFKKSEEEADASISEEEEESEPPTKK
IKMDNNSVGGSGTSLSA*
>G396 (1..957)
ATGGGGGAAAGAGATGATGGGTTGGGTTTGTAGTCTAAGCTTGGGAAATAGTCAACAAAAA
GAACCATCTCTGAGGTTGAATCTTATGCCGTTGACAACCTTCTTCTTCTTCTTCTGTTT
CAACACATGCACAATCAGAATAACAATAGCCATCCCCAGAAGATTCAATAACATCTCTTGG
ACTCATCTGTTTCAATCTTCTGGGATTAAACGTACAACCTGCAGAGAGAACTCCGACGCC
GGGTCAATTTCTAAGAGGTTTCAACGTGAACAGAGCTCAGTCTTCCGGTGGCGGTAGTGGAC
TTGGAAGAAGAAGCCGCCGTCGTCTCGTCTCCTCAACAGCGCCGTTTTCGAGTCTGAGTGGA
AATAAAAGGATCTTGCGGTGGCGAGAGGAGATGAAAACGAGGCGGAGAGAGCTTCT
TGCTCAGCGGAGGGGGAAGCGGTGGTAGCGACGATGAAGACGGCGGAAACGGCGACGGA
TCAAGGAAGAACTACGGTTATCGAAGGATCAAGCTCTTGTCTCGAGGAGACTTTTAAA
GAACATAGCACTCTTAATCCGAAGCAAAAGCTGGCTCTAGCAAAACAGTTGAATCTAAGG
GCAAGACAAGTTGAAGTGTGGTTTTCAGAACCGTAGGGCAAGGACGAAGCTGAAACAAACG
GAGGTTGATTGTGAGTATTTAAAGAGATGTTGCGATAATCTGACCGAGGAGAATCGACGG
CTGCAGAAAGAAGTGTGCGAGCTGAGGGCGTTGAAGTTGTCTCCACATCTCTACATGCAC
ATGACTCCTCTACTACTCTCACCATGTGCCCTTCTTGGCAACGTGTCTCCTCTCTGCC
GCCACTGTGACCGTGTCTCTTCCACTACTACTCTCTACGTTGGTGGGGCGGCCAAGT
CCACAGCGATTAACTCCTTGGACTGCTATTTCTCTCCAGCAAAATCAGGTCGCTAG
>G396 Amino Acid Sequence (domain in AA coordinates: 159-220)
MGERDDGLGLSLSLGNSQQKEPSLRLNLMPLTSSSSSFQHMHNQNNNSHPQKIHNISW

>G431 (1..1149)

[illegible]

ATGGAGATGGGTTCCAACACTCGGGTCCGGGTTCATGGTCCGGGTTCAGGCAGAGTCCGGGTGGT
TCCTCCACTGAGTCATCCTCTTTTCAGTGGAGGGCTCATGTTTGGCCAGAAGACTACTTTC
GAGGACGGTGGTGGTGGATCCGGGTCTTCTTCCTCAGTGGTGGTTCGTTCAAACAGACAGCTGTC
CGTGGAGGCGGGTTCGGGTTCAGTCCGGGTTCAGATACCAAGGTGCCAAGTGGAAAGGTTGTGGG
ATGGATCTAACCAATGCAAAGGTTATTACTCGAGACACCGAGTTTGTGGAGTGCACCTCT
AAAACACCTAAAGTCACTGTGGCTGGTATCGAACAGAGGTTTTGTCAACAGTGCAGCAGG
TTTCATCAGCTTCCGGAATTTGACCTAGAGAAAAGGAGTTGCCGCAGGAGACTCGCTGGT
CATAATGAGCGACGAAGGAAGCCACAGCCTGCGTCTCTCTCTGTGTTAGCTTCTCGTTAC
GGGAGGATCGCACCTTCGCTTTACGAAAATGGTGTGCTGGAATGAATGGAAGCTTTCTTT
GGGAACCAAGAGATAGGATGGCCAAGTTCAGAACATTGGATACAAGAGTGATGAGGCGG
CCAGTGTCTGTCACCGTTCATGGCAGATCAATCCAATGAATGTATTTAGTCAAGGTTTCAGTT
CGTGGAGGAGGCAACAGCTTCTCATCTCCAGAGATTATGGACACTAAACTAGAGAGCTAC
AAGGGAATTTGGCGACTCAAACGTGCTCTCTCTCTCTCTGTCAAATCCACATCAACCACAT
GACAACAACAACAACAACAACAACAACAACAACAACAACAATACATGGCGAGCTTCT
TCAGGTTTTTGGCCCGATGACGGTTTCAATAGGCTCAACCACCACCTGCACCTAGCCAGCAT
CAGTATCTGAACCCGCCTTGGGTATTCAAGGACATAAGTATGATATGTCTCCTGTTTTG
AATTTAGGTCGATACACCGAGCCAGATAATTGTGCAGATAAGTAGTGGCAGCGCAATGGGT
GAGTTTCAGTTATCTGATCACCATCATCAAAGTAGGAGACAGTACATGGAAGATGAGAAC
ACAAGGGCTTATGACTCTTCTCTCTCACCATACCAACCTGGTCTCTCTGA

129

FHQLPEFDLEKRSRRRLAGHNERRRKQPASLSVLASRYGRIAPSLYENGDA GMNGSFL
GNQEIGWPSSRTLDTRVMRRPVSSPSWQINPMNVFSQGSVGGGTSFSSPEIMDTKLESY
KGIGDSNCALSLLSNP HQPHDNNNNNNNNNNNNNTWRASSGFGPMTVTMAQPPPAPSQH
QYLNPPWVFKDNDNDMS PVLNLGRYTEPDNCQISSGTAMGEFELSDHHHQSR RQYMEDEN
TRAYDSSSHHTNWSL*

>G546 (1..588)

atgactcgaccgtcaagattacttgagacggcgccaccaccacaaccgtcggaggag
atgatcgacgcggaatccgacatggtggtgatcttgtcggctcttcttgcgctcttatac
tgcggttgctggtctcgccgcgctcgtagatgcgcttggtccggcggtttacagccgga
ggagattcgccgtcaccgaacaaaggcttgaaaaagaaagctcttcagtctcttccaaga
tccactttcaccgcgcggaatcaacctccggcgccgcgctgaagaggagactcgacg
gaatgtgctatttgctcactgacttcgccgacggtgaagaaataagagtgtcttctctt
tgtggtcattcttccacgtggagtgtattgacaaatggctagtttctaggtcttcttgt
ccttcttgtcgcaggattcttacgcgggtgagatgtgaccggtgtggtcatgttctacg
gcggagatgaaagatcaagctcatcgatcaacatcaccaacactcttctactaccatt
cctacgtttcttctcttaa

>G546 Amino Acid Sequence (domain in AA coordinates:114-155)
MTRPSRLLETAAPPPQSEEMIAESDMVVILSALLCALICVAGLA AVRCALRRFTAG
GDSPSPNKGLKKKALQSLPRSTFTAESTSGAAAEEDSTECAICLTD FADGEBIRVLPL
CGHSFHVECIDKWLVSRS SPCRRILTPVRCDRCGHASTAEMKDQ AHRHQHQSSTTI
PTFLP*

>G551 (1..708)

ATGGAGTGGTCAACAACGAGCAACGTAGAAAACGTGAGAGTAGCTTTCATGCCACCGCCA
TGGCCGGAGTCTAGTTCTTTAACTCGCTCCACAGCTTCAACTTTGATCCTTACGCAGGA
AATTCATATACGCCTGGCGATACACAAACCGACCGGTTATCTCTGTACCGGAATCAGAA
AAGATCATGAATGCGTACCGATTTCCGAACAACAATGAGATGATAAAAAAGAAGAGA
CTAACGAGTGGACAATTAGCTTCACTTGAGCGAAGTTTTCAAGAAGAGATCAAATTAGAT
TCAGACAGGAAGGTGAAGCTGTGAGAGAGCTCGGTCTGCAGCCACGTGAGATAGCAGTT
TGGTTCCAAAACCGCGGTGCACGGTGGAAGGCGAAGCAGCTTGAGCAGTTGTACGACTCG
CTTAGACAAGAGTACGACGTCGTTTCTAGGGAGAAACAAATGTTACACGATGAGGTGAAG
AAGCTGAGAGCTTTACTAAGAGACCAGGGTTTGATCAAGAAGCAAATCTCTGCCGGGACC
ATCAAAGTTTCCGGTGAGGAAGACACGGTGGAGATTTATCGGTGGTGGTAGCTCATCCA
AGAACGGAGAATGAACGCAAAATCAAATCACCGAGGGAATCAAGTTTACGGTCAATAC
AACAAATCCGATGCTGGTTGCTTCTCTGGCTGGCCGTCATACCCCTGA

>G551 Amino Acid Sequence (conserved domain in AA coordinates:73-133)
MEWSTTSNVENVRVAFMPPWPPESSSFNSLHSFNFDPYAGNSYTPGDTQTGPVISVPESE
KIMNAYRFPNNNNEMIKKRLTSGQLASLERSFQEEIKLSDRKVKLSRELGLQPRQIAV
WFQNRRLRWKAKQLEQLYDILRQEYDVVSREKQMLHDEVKKLRALLRDQGLIKQISAGT
IKVSGEEDTVEISSVVVAHPR TENMNANQITGGNQVYQYNNPMLVASSGWPSYP*

>G578 (1..978)

ATGCATAGTTTGAATGAAACAGTAATTCCTGATGTTGATTACATGCAGTCTGATAGAGGG
CATATGCATGCTGCTGCCTCTGATTCCAGTGATCGATCAAAGGATAAGTTGGATCAAAAG
ACCCCTCGTAGGCTTGCTCAAAATCGTGAGGCAGCAAGAAAAGCAGATTGAGGAAGAAG
GCGTATGTTT CAGCAGCTGGAAGATAGTCGATTAAAGCTGACTCAAGTTGAGCAGGAGCTG
CAAAGAGCAAGACAGCAGGGAGTTTTCATCTCAAGTT CAGGAGACCAAGCTCATTCTACT
GGTGGCAATGTTGGGGCTTTGGCATTGATGCAGAACACTCACGATGGCTTGAAGAAAAG
AACAGGCAAATGAACGAGCTGAGATCTGCCCTGAATGCTCATGCAGGTGATACTGAGCTC
CGGATAATTGTGGATGGAGTGATGGCTCACTATGAGGAGCTTTTCAGGATTAAGAGCAAT
GCATCTAAGAATGATGCTCTCCACTTGTTATCTGGAATGTGGAACCAACAGCTGAGCGA
TGTTTCTTGTGGCTTGCGGGTTCCCGTCATCCGAACCTTCTCAAGCTTCTTGCGAATCAG
CTAGAGCCCATGACAGAACGACAGGTAATGGGCATCAATAGCTTGACAGACGTCGCAG
CAGGCAGAAGATGCTTTATCTCAAGGGATGGAGAGTTTACAGCAATCCCTAGCTGATACT
TTATCCAGTGGAACCTTTGGTTCCAGTTCATCGGATAATGTCGCGAGCTACATGGGT CAG
ATGGCCATGGCAATGGGCAAGTTAGGCACCCTCGAAGGATTCATACGCCAGGCTGATAAC
TTGAGGCTGCAAACTACAACAGATGCTTCGAGTATTAACAACACGTCAGTCAGCTCGT
GCTCTTCTTGCTATACAGATTATTCATCTCGATTACGTGCTCTTAGTTCCTTGTGGCTT
GCCCCGCCAAGAGAGTGA

>G578 Amino Acid Sequence (domain in AA coordinates 36-96)
MHSLNETVIPVDVDMQSDRGHMHAAASDSSDRSKDKLDQKTLRRLAQNREAARKSRLRKK
AYVQQLSDSRLKLTQVEQLQRRARQQGVFISSSGDQAHSTGGNGGALAFDAEHSRWLEEK
NRQMNELRSALNAHAGDTELRIIVDGVMAHYEELFRIKSNASKNDVFHLLSGMWKTPAER
CFLWLGGFPSSSELLKLLANQLEPMTERRQVMGINSLQQTQQAEADALSQGMESLQQSLADT
LSSGTLGSSSSDNVASYMGQMAMAMGKLGTLGFIHQADNLRQLQQLRVLTTTRQSAR
ALLAIHDYSSRLRALSSLWLARPRE*

>G596 (168..1121)

TAATTTCTCTACTTCAGATTTTTTCTCCTTAGATTAATTTAATTGAGTTATTGTACATC
CCTCAAGCTAAGATTCTGGTTTTGTGAGTTGAGTGGATGAGAAGAGGAGAGATTAACTAA
ATTAGGGTTTCAATTGTTTACTTTTTGTTTGTCTTTTATATCAAGTAATGGATCAGGTCT
CTCGCTCTCTTCTCCACCTTTTCTCTCAAGAGATCTCCATCTTCACCACACCATCAAT
TCCAGCATCAGCAGCAGCAGCAACAGAATCACGGCCACGATATAGACCAGCACCAGAA
TCGGTGGGCTAAAACGTGACCGAGATGCTGATATCGATCCCAACGAGCACTCTTCAGCCG
GAAAAGATCAAAAGTACTCTCGCTCCGGTGGAGAAAAGCGCGCGGAGGAGGAGAGATA
ATCACATCAGGAGAAGGCCACGTGGCAGACCAGCGGGATCTAAGAACAACCAAAACCGC
CAATCATCATCACTCGAGACAGCGCAAACGCTCTCAAATCTCATGTCTGGAAGTAGCAA
ACGGATGTGACGTCTGGAAGGTGTCACCGTCTTCGCTCGCCGTCGCCAACGTGGCATCT
GCGTTTTGAGCGGAAACGGCGCGGTTACCAACGTTACCATAAGACAACAGCTTCAGTAC
CTGGTGGTGGCTCATCTGTCGTTAACTTACACGGACGTTTTGAGATTCTTTCTCTCTCGG
GATCATTCTCTCTCTCCGGCTCCACAGCTGCGTCAGGTCTAACGATTTACTTAGCCG
GTGGTCAGGGACAGGTTGTTGGAGGAAGCGTGGTGGTCCACTCATGGCTTCAGGACCTG
TAGTGATTATGGCAGCTTCGTTTGGAAAACGCTGCGTATGAGAGACTGCCGTTGGAGGAAG
ACGATCAAGAAGAGCAAACAGCTGGAGCGGTTGCTAATAATATCGATGGAAACGCAACAA
TGGGTGGTGGAAACGCAAAACGCAAACTCAGACGCAGCAGCAACAGCAACACAGTTGATGC
AAGATCCGACGTCGTTTATACAAGGGTTGCCCTCCGAATCTTATGAATCTGTTCAATTGC
CAGCTGAAGCTTATTGGGGAACCTCCGAGACCATCTTCTAAATCGCGAAGAAAAACAAG
TTAGATACGTTTCGTTGTTTTTAATTTATAATCTCTCTCTGTCAAGTTTAAATTTCTTT
TTCTTCTCTTTGTTTTCTAAAGATAATTGTAGTCTTTGACGAAGATTCGTGGTACGTAT
GAATCGAAGAGAATCGTTTTGGTTCATGGGATTGCTCGATCTATTAGGTTTGAGAGGGGGT
TTGTGTTTTGCGTTGACTAGCAGATTATAAAATTGTTGATTTTCGAGTTTATTTTCAT
GTGTTGGTGATAAA

>G596 Amino Acid Sequence (domain in AA coordinates: 89-96)

MDQVSRSRLPPPFLSRDLHLHPHHQFQHQQQQQQNHGHDIDQHRIGGLKRDRDADIDPNE
HSSAGKDQSTPGSGGESGGGGGDNHITRRPRGRPAGSKNPKPPIIITRDSANALKSHV
MEVANGCDVMESVTVFARRRQRGICVLSGNGAVTNVTIRQPASVPGGSSVVLNHRFEI
LSLSGSFLPPPAPPAASGLTIYLAGGQGVVGGSSVVGPLMASGPVIVMAASFVGNAYERL
PLEEDDQEEQTAGAVANNIDGNATMGGGTQTQTQTQQQQQQQLMQDPTSFQGLPPNLMN
SVQLPAEAYWGTTPRPSF*

>G617 (59..1141)

CAGATCTGTTCTTTACACCAAATTGAGTACTGAAGATCTTGTGAGTGAATTAAAGAGAT
GAGATCAGGAGAATGTGATGAAGAGGAGATTCAAGCAAAGCAAGAAAGAGATCAAAATCA
AAATCATCAAGTAAACTTAAACCACATGTTGCAACAACAACAGCCGAGTTCGGTATCATC
TTCAAGGCAATGGACTTCAGCTTTTAGGAATCCAAGAATCGTTCGAGTCTCAAGAACATT
CGGTGGCAAAGACAGACACAGCAAAGTATGTACAGTCCGTGGTCTTCGAGACCGGAGGAT
AAGGTTGTCCGTACCTACAGCTATTCAACTCTACGACCTTCAAGATCGATTAGGGCTGAG
TCAGCCAAAGCAAAGTCATTGATTGGTTACTCGAAGCAGCAAAAGATGACGTAGACAAGCT
ACCTCCTTACAAATCCCACATGGATTAAACAGATGTATCCAATCTCATCTTCGAAA
CTCCGGGTTTGGAGAATCTCCATCTTCAACTACATCAACAACGTTTCCAGGAACCAATCT
CGGGTTCTTGGAAAATTGGGATCTTGGTGGTCTTCAAGAACAAGAGCAAGATTAACCGA
TACAACTACGACCCAAAGAGAAAGTTTGTATCTTGATAAAGGAAAATGGATCAAAAACGA
CGAGAATAGTAATCAAGATCATCAAGGGTTTAAACCAATCATCAACAACAATTTCTCT
GACCAATCCGTACAACAACACTTCAGCTTATTACAACCTTGGACATCTTCAACAATCGTT
AGACCAATCTGGTAATAACGTTACTGTGCGCAATATCTAATGTTGCTGCTAATAATAACAA
TAATCTCAATTTGCATCTCTCTCTCGTCTGCCGGAGATGGATCTCAGCTTTTTTTCGG
TCCTACTCTCCGCAATGAGCTCTCTATTCCCGACATACCTTCGTTTCTTGGAGCTTC
TCATCATCATCATGTCTGTCGATGGAGCCGGTCATCTTCAGCTCTTTAGCTCGAATTCAAA

TACCGCATCGCAGCAACACATGATGCCGGTAATACGAGTTTGATTAGACCATTTCATCA
TTTGATGAGCTCGAATCATGATACGGATCATCATAGTAGCGATAATGAATCAGATTCTTG
AATGATTTTATATATCTACACTATACATTGAAAATGTTATATGTATACGTATTCTTCTAT
ATTTTGATATATATGCGTATTGTTGGATTGGTTTATGTATCT

>G617 Amino Acid Sequence (domain in AA coordinates: 64-118)
MRSGECDDEEIIQAKQERDQNHQVNLNHLQQQPSSVSSSRQWTSAFRNPRIVRVSR
FGGKDRHSKVCTVRGLRDRIRLSVPTAIQLYDLQDRLGLSQPSKVIDWLLEAAKDDVDK
LPPLQFPHGFNQMYPNLIFGNSGFGESPSSTTSTTFPGTNLGFLENWDLGGSSRTRARLT
DTTTTQRESFDLDKQKWIKNDENSNQDHQGFNTNHQQQFPLTNPYNNTSAYYNLGHLLQOS
LDQSGNNVTVAISNVAANNNNNLNLHPPSSSAGDGSQLFPGTPPAMSSLFPTYPSFLGA
SHHHHVVDGAGHLQLFSSNSNTASQQHMPGNTSLIRPFHMLSSNHDTDHSSDNESDS
*

>G620 (40..666)
GAATTGAACCTGGACCAGCACAGCAACAACCCCAACCCCAATGACCAGCTCAGTCATAGTA
GCCGGCGCCGGTGACAAGAACAATGGTATCGTGGTCCAGCAGCAACCACCATGTGTGGCT
CGTGAGCAAGACCAATACATGCCAATCGCAAACGTCATAAGAATCATGCGTAAACCTTA
CCGTCTCAGCCAAAATCTCTGACGACGCCAAAGAAACGATTCAAGAATGTGTCTCCGAG
TACATCAGCTTCGTGACCGGTGAAGCCAACGAGCGTTGCCAACGTGAGCAACGTAAGACC
ATACTGCTGAAGATATCTTTGGGCTATGAGCAAGCTTGGGTTTCGATAACTACGTGGAC
CCCTCACCCTGTTCATTAAACCGGTACCGTGAGATAGAGACCGATCGTGGTTCTGCACTT
AGAGGTGAGCCACCGTCGTTGAGACAAACCTATGGAGGAAATGGTATTGGGTTTCACGGC
CCATCTCATGGCTACCTCCTCCGGTCTTATGGTTATGGTATGTTGGACCAATCCATG
GTTATGGGAGGTGGTCCGTACTACCAAAACGGGTCTGCGGTCAAGATGAATCCAGTGTT
GGTGGTGGCTCTCGTCTTCCATTAAACGGAATGCCGGCTTTTGACCATTATGGTCAGTAT
AAGTGAAGAAGGAGTTATTCTTCTTATATCTATTCAAACATGTGTTTCGATAGAT
ATTTTATTTTATGTCTTATCAATAACATTTCTATATAATGTTGCTTCTTTAAGGAAAAG
TGTTGTATGTCAATACTTTATGAGAACTGATTTATATATGCAAAAT

>G620 Amino Acid Sequence (domain in AA coordinates: 20-118)
MTSSVIVAGAGDKNNGIVVQQPPCVAREQDQYMPIANVIRIMRKTLP SHAKISDDAKET
IQECVSEYISFVTGEANERCQREQRKTITAEDILWAMSKLGFNDYVDPLTVFINRYREIE
TDRGSALRGEPPSLRQTYGGNGIGFPHGPHSHGLPPPGPYGYGMLDQSMVMGGGRYYQNGSS
GQDESSVGGSSSSSINGMPAFDHYGQYK*

>G625 (151..1137)
AATCGACCAATTCACAACGATGACATTCAAACACTCTTCAGTTTCCCTTCCTTCTTGATT
GTCTCTCCACTATTTTCTCAATTTCTTTAATCTCTCTCTTCTCTCTCTACTTCTCT
TCCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
AACCATCTGGAAGATAATAACCAAAACCTAACCCATAATAATCTCAATCCGATTCCACC
ACCGACTCATCAACTTCCTCCGCTCAACGCAACGCAAGGCAAGGTGGTCCGGACAAC
TCCAAGTTCCGTTACCGTGGCGTTCGACAAAGAAGCTGGGGCAATGGGTCCGGCAGATC
CGAGAGCCACGTAAGCGCACTCGCAAGTGGCTTGGTACTTTTCGCAACCGCCGAAGACGCC
GCACGTGCCTACGACCGGCTGCCGTTTACCTATACGGGTCACGTGCTCAGCTCAACTTA
ACCCCTTCGTCTCCTTCTCCGCTCTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT
CCTTCCACCTCCTTCTTCTTCCACTCAAACCTAAGACCTCTCCTCCCTCGCCCCGCGCC
GCCACCGTAGGAGGAGGACCAACTTTGGTCCGTACGGTATCCCTTTTAAACAACAACATC
TTCCTTAATGGTGGGACCTCTATGTTATGCCCTAGTTATGGTTTTTTCCTCAACAACAA
CAACAACAAATCAGATGGTCCAGATGGGACAATCCAACACCAACAGTATCAGAATCTT
CATCTAATACTAACAATAACAAGATTTCTGACATCGAGTCACTGATGTTCCGGTAACT
AATTCGACTTCGTTTCATCATGAGGTGGCGTTAGGGCAGGAACAAGGAGGAAGTGGGTGT
AATAATAATAGTTCGATGGAGGATTTGAACCTCTAGCTGGTTCGGTGGGTTCGAGTCTA
TCAATAACTCATCCACCGCGTTGGTTGATCCGGTATGTTCTATGGGTCTGGATCCGGGT
TATATGGTTGGAGATGGATCTTCGACCATTGGCCCTTTTGGAGGAGAAGAAGAATATAGT
CATAATTGGGGGAGTATTTGGGATTTTATGATCCCATCTGGGGGAATTCTATTAAATTT
GTTTTTGTGGAAGATCATATTATATACGATGAGCATCCCTAAGGTCCGTCAAGAGCATTG
GAGATTCAATGTTGAGAGGAATCAAAGAGATTGCATTCTATGAGGAGCTCTGCATGCAAA
ATTTTGGAGGATTTTTTACTACCTATAGAGATAAATAAGAGGGTATTTTTATTATTTTT
TTGAAGATTTTTATTTCAGGAATTCGTAAGAGATTACGGTTCCAATAAAGTATGTA
TATGTGGAAGAGAATCGGAGGAGATGGTGGAAAGTTGTATGGGAATTTTATTGGTTCAAC

>G658 (17..757)

>G658 Amino Acid Sequence (domain in AA coordinates: 2-105)

>G716 (271..2079)

133

ACTCCTCCTCCTCCAACGAATTGTAGCTATAGGTTGTTTGGATTGTGATCTCACAAGCAAT
TCTCCTGCCTCAATCCCTCAAGACAAGCAACCGATGGATACTTGTGGAGCTGCCAAGTGT
CAAGAACCCATCACTCCAACCTCAATGAGTGAGCAGAAGAAGCAACAAACATCAAGAAGT
CGAACTAAAGTGCAAATGCAAGGCATTGCGGTTGGTCTGCGGTTGATTAAACACTGTTG
AAATCTTACGATGAACTGATTGATGAGCTTGAGGAGATGTTTGAGATTCAAGGACAGCTT
CTTGCCCGAGACAAATGGATCGTTGTCTTCACTGATGATGAAGGAGATATGATGCTTGCT
GGTGATGATCCGTGGAATGAGTTTTCGAAGATGGCAAAGAAGATATTTATATATTCGAGC
GATGAGGTTAAGAAAATGACAACGAAACTGAAGATTTCTTCGTCGTTAGAGAATGAGGAA
TATGGTAATGAATCATTTCGAAAATCGTAGTAGGGGGTGAGAGTTTTAGCTGTTAATTAAG
GTTAATTTCGGCGACGTCGTTTTAGTGCCTAAGTGTCTAAAGACTTTTTTTTTTAGTCTGTG
TATATAAAGTCTTGTCTCTTTTTTCATGTCAATTTTTCAAGTTGGCGATTTAATATTTG
GTTTTGGGACAGTGGTTGATGGGGCGGTTTTACATTTTTTATGTGTATGTACTTGTTC
AAACCATTCAATTTTCAAA

>G716 Amino Acid Sequence (domain in AA coordinates: 24-355)
MASVEGDDDFGSSSSRSYQDQLYTELWKVCAGPLVEVPRAQERVVFYFPQGHMEQLVASTN
QGINSEEIIPVFDLPPKILCRVLDVTLKAEHETDEVYAQITLQPEEDQSEPTSLDPPIVGP
TKQEFHSFVKILTASDSTHGGFSVLKHKATECLPSLDMTQATPTQELVTRDLHGFEWRF
KHIFRGQPRRHLLTTGWSTFVSSKRLVAGDAFVFLRGENGDLRVGVRRLLARHQSTMPTSV
ISSQSMHLGVLATASHAVRTTIFVVFYKPRISQFIVGVNKYMEAIAKHGFSLGTRFRMR
EGEESPERIFTGTIVGSGDLSSQWPASKWRSLLQVQWDEPTTVQRPDKVSPWEIEPFLATS
PISTPAQQPQSKCKRSRPIEPSVKTPAPPSFLYSLPQSQDSINASLKLFDPSLERISGG
YSSNNSFKPETPPPTNCSYRLFGLTNSNPAPIQDKQPMDCGAACQEPITPTSMS
EQKKQQTSSRSRTKVQMGIAGRAVDLTLLKSYDELIDELEEMFEIQGQLLARDKWIVVF
TDDEGDMMLAGDDPWNEFCMAKKIFIYSSDEVKMTTKLKISSSLENEEYGNESFENRS
RG*

>G725 (46..1122)
CCTCTTTTCAGAGAGAGAAAGAGAGTCAGAGAGAGAGAGAGAGAGAATGTTCCATGCTAAG
AAACCTTCAAGTATGAATGGTTCATATGAGAACAGAGCTATGTGCGTTCAAGGCGATTCA
GGCCTTGTCTCACCACCGACCCTAAACCGCGTTTTCGTTGGACCGTCGAACCTCCACGAG
CGTTTTGTGGACGCGCTCGCTCAGCTCGGCGCCCCGACAAAGCGACCCCAAGACGATT
ATGAGAGTTATGGTGTGAAGGGTCTTACTCTTTACCACCTAAAGAGCCATCTTCAGAAA
TTCAGGCTTGGAAGCAGCCGCACAAGGAGTACGGAGATCACTCCACAAAGGAAGGTTCA
AGAGCTTCGCCATGGATATTACGCGCAACGTAGCTTCTTCTTCTGGCATGATGAGTCGC
AACATGAATGAGATGCAAATGGAAGTGCAGAGAAGGTTGCATGAACAGCTAGAGGTGCAA
AGACATCTGCAACTGAGGATTGAAGCACAAGGAAAGTACATGCAATCTATCTTGGAGAGA
GCTTGCCAAACCTAGCCGGTGAGAACATGGCAGCCGCCACCGCAGCAGCCGCCCTCGGA
GGAGGATACAAGGGTAATCTGGGAAGTTCGAGTCTTTCAGCAGCGGTGGGCCCCACCTCCT
CATCCTCTTAGTTTCCCGCGGTTTCAAGACCTAAACATCTATGGAACACAAACCGACCAA
GTCCTCGACCATCACAACCTCCATCATCAAAACATAGAGAACCATTTACCGGTAACAAT
GCTGCAGACACCAACATTTACTTGGGGAAGAAGCGACCTAATCTAATTTTGGTAACGAT
GTAAGGAAAGGACTATTGATGTGGTCTGATCAAGATCACGATCTTTCGCAAAACCAATCG
ATCGATGATGAGCATAGAATTACAGATACAGATGGCTACACATGTCTCCACGGATTTGGAT
TCTTTGTGCGAGATCTACGAAAGGAAATCAGGTTTATCAGGTGATGAAGGGAATAATGGT
GGGAAATTACTGGAAGGCCATCGCCTAGGAGATCACCATTTAGTCTTATGATGAACCTT
AATGGTGGATTAATAACAAGGAAGAACTCGCCATTTGGGTGATACAATTTATTAATTTTT
ATCTATGAGTGATGCATGGGAATGTAAGAACGAGATATATATGTTTTGTCAATTGTGAGTT
TGACGTAGGGTTTAGAGAAAA

>G725 Amino Acid Sequence (domain in AA coordinates: 39-87)
MFHAKKPSSMNGSYENRAMCVQDGLVLTDPKPLRWTVELHERFVDAVAQLGGPDKA
TPKTIMRVMGVKGLTLYHLKSHLQKFRLLGKQPHKEYGDHSTKEGSRASAMDIQRNVASSS
GMMSRNMNEMQMEVQRRLEHQLEVRHLQLRIEAGQKYMQSILERACQTLAGENMAAATA
AAAVGGGYKGNLSSSLAAVGGPPHPLSPFPFQDLNIYGNTTDQVLDHNFHHQNIENH
FTGNNAADTNIYLGKKRPNPNFGNDVRKGLLMWSDQDHDLSANQSIDDEHRIQIQMATHV
STDLDLSLEIYERKSGLSGDEGNNGKLLERPSRRSPLSPMMNPNGGLIQGRNSPFG*
>G727 (43..1977)
CTTCTTCTCCTTCTCTGATCGTTCTGTTTCTGGACGAGAGAGATGGTAAATCCGGGTCAC
GGAAGAGGACCCGATTCCGGGTACTGCTGCTGGTGGGTCAAACCTCCGACCCGTTTCTGCG

AATCTTCGAGTTCTTGTCTGATGATGATCCAACCTTGTCTCATGATCTTAGAGAGGATG
CTTATGACTTGTCTCTACAGAGAGCAGAGAGCGCATTGTCTCTGCTTCGGAAGAACAAG
AATGGTTTGTGATATTGTCATTAGTGATGTTTCATATGCCTGACATGGATGGTTTCAAGCTC
CTTGAACACGTTGGTTTAGAGATGGATTTACCTGTTATCAATCTGAATGTTTTGAAACCT
TTGGTTATAGTGATGCTCTCGGATGATTCTGAAGAGCGTTGTGTTGAAAGGAGTGACTCAC
GGTGAGTTGATTACCTCATCAAAACCGGTACGTATTGAGGCTTTGAAGAATATATGGCAA
CATGTGGTGCAGGAAGAAGCGTAACGAGTGGAATGTTTCTGAACATTCTGGAGGAAGTATT
GAAGATACTGGCGGTGACAGGGACAGGCAGCAGCAGCATAGGGAGGATGCTGATAACAAC
TCGTCTTCAGTTAATGAAGGGAAACGGGAGGAGCTCGAGGAAGCGGAAGGAAGAGGAAGTA
GATGATCAAGGGGATGATAAGGAAGACTCATCGAGTTTAAAGAAACCACGCGTGGTTTGG
TCTGTTGAATGTCATCAGCAGTTTGTGCTGCTGTGAATCAGCTAGGCGTTGACAGTGAG
TTAAAACTTGCTTGCTTATGCAATTTGTGTGTGTCGATTGTAACATTGTGGAATTCCAG
AAGTATCGGATATATCTGAGACGGCTTGGAGGAGTATCGCAACACCAAGGAAATATGAAC
CATTCGTTTATGACTGGTCAAGATCAGAGTTTGGACCTCTTCTTCGTTGAATGGATTT
GATCTTCAATCTTTAGCTGTTACTGGTCAGCTCCCTCCTCAGAGCCTTGACAGCTTCAA
GCAGCTGCTTGGCCGGCTTACACTCGCTAAACCAGGGATGTCGGTTTCTCCCTTGTA
GATCAGAGAAGCATCTTCACTTTGAAAACCCAAAAATAAGATTTGGAGACGGACATGGT
CAGACGATGAACAATGGAAATTTGCTTCATGGTGTCCCAACGGGTAGTCACATGCGTCTG
CGTCTGGACAGAATGTTTCAGAGCAGCGGAATGATGTTGCCAGTAGCAGACCAGCTACCT
CGAGGAGGACCATCGATGCTACCATCCCTCGGGCAACAGCCGATATTGTCAAGCAGCGTT
TCAAGAAGAAGCGATCTCACTGGTGCCTGGCGGTTAGAAACAGTATCCCCGAGACCAAC
AGCAGAGTGTTACCAACTACTCACTCGGTCTTCAATAACTTCCCCGCGGATCTACCTCGC
AGCAGCTTCCCGTTGGCAAGTGCCCCAGGGATTTAGTTCCAGTATCAGTTTCTTACCAA
GAAGAGGTCAACAGCTCGGATGCAAAAGGAGGTTTCATCAGCTGCTACTGCTGGATTTGGT
AACCCAAGCTACGACATATTTAACGATTTTCCGCAGCACCAACAGCACACAAGAATC
AGCAATAAACTAAACGATTGGGATCTGCGGAATATGGGATTGGTCTTCAGTTCCAATCAG
GACGCAGCAACTGCAACCGCAACCGCAGCATTTCCTCACTTCGGAAGCATACTCTTCGTCT
TCTACGCAGAGAAAAAGACGGGAAACGGACGCAACAGTTGTGGGTGAGCATGGGCAGAAC
CTGCAGTCACCGAGCCGGAATCTGTATCATCTGAACCACGTTTTTATGGACGGTGGTTCA
GTCAGAGTGAAGTCAGAAAGAGTGGCGGAGACAGTGACTTGTCTCCAGCAAATACATTG
TTTCAGGAGCAGTATAATCAAGAAGATCTGATGAGCGCATTTCTCAAACAGGTTTGATTA
TTACTCGAATACAGTGCATCTAAAAAC

>G727 Amino Acid Sequence (domain in AA coordinates: 226-269)
MVNPGHGRGPDSTAGGSNSDPFANLRVLVDDDPCLMILRMLMTCLYREQRAHCL
CFGRKNGFDIVISDVHMPDMDGFKLLEHVGLEMDLPVINLNLKPLVIVMSADDSKSVV
LKGVTGAVDYLKIPVRIEALKNIWQHVVRRKRNEWNVSEHSGGSIEDTGGDRDRQQHR
EDADNNSSSVNEGNRRSRKRKEEVDDQGDDESSSLKPRVWVSVLHQFVAAVNQ
LGVDSELKTCLLMHLCVSIGNIVEFQKYRIYLRRLGGVSVQHQNMMNHSFMTGQDQSFGL
SSLNGFDLQSLAVTGQLPPQSLAQLQAAGLGRPTLAKPGMSVSPVLDQRSIFNFENPKIR
FGDGHGQTMNNGNLLHGVTGSHMRLRPGQNVQSSGMLPVADQLPRGGPSMLPSLGQQP
ILSSSVRRSDLTGALAVRNSIPETNSRVLPTTHSVFNNFPADLPRSSFPLASAPGISVP
VSVSYQEEVNSSDAKGGSSAATAGFGNPSYDIFNDFPOHQHKNKINSKLNWDLRNMGL
VFSSNQDAATATATAAFSTSEAYSSSSTQRKRRETDATVVGEHGQNLQSPSRNLYHLNHV
FMDGGSVRVKSERVAETVTCPPANTLFEHQYNQEDLMSAFLKQV*

>G740 (25..924)

CTTCTTCAACTTTTTTTTAAACGATGGCTTCAGAGGATCAATCGGCGGCGAGATCTACC
GGGAAGGTGAAGTTCACGCTTCTAAAGGCTATGGTTTCATTACTCCTGACGATGGC
AGCGTAGAGCTTTTGTTCATCAATCTTCAATTGTCTCCGAAGGTTACCGGAGTTTAAAC
GTCCGGGATGCGGTTGAGTTGCTATTACTCAGGGAAGCGACGGTAAGACTAAAGCCGTC
AATGTTACTGCTCCTGTTGGTGGTTCTCTCAAGAAGGAGAATAACTCTCGTGGTAACGGT
GCTAGGCGCGGCGGCGGTGGAAGCGGTTGCTACAATTGCGGTGAGTTAGGTCATATCTCT
AAAGATTGTGGTATTGTTGGCGGCGGCGGAGGTGGTGAACGTAGATCTAGAGGAGGAGAA
GGTTGTTACAATTGTGGTGATACTGGTCACCTCGCTAGGGATTGTACTTCAGCTGGAAAC
GGTGACCAACGTGGAGCCACCAAGGTGGAAACGATGGTTGCTACACTTGCAGGTGATGTT
GGTCAGTGGCTAGGGATTGTACTCAGAAATCAGTTGGAAACGGAGACCAACGTGGAGCG
GTCAAAGGTGGAAACGATGGTTGCTACACTTGTGGTGATGTTGGTCACTTTGCTAGGGAT
TGTAATCAGAAGGTTGCTGCCGGAACGTCAGAAGCGGTGGTGGTGGTAGTGGAACCTGT

TATTCATGCGGTGGAGTTGGTCACATTGCAAGAGATTGTGCGACTAAGAGACAGCCTTCT
CGTGGGTGTTACCAAGTGTGGTGGTCTGGTCACCTGGCTCGTGATTGTGACCAGAGAGGA
AGCGGTGGAGGAGGTAATGATAATGCGTGCTACAAGTGTGGTAAGGAAGGTCACCTTTGCA
AGGGAATGTTCTTCTGTAGCTTAATCGATTTCCTAATCAACAAAACAAAAACAAGAAT
GAAATTGAATCGAGTTATATAGTTTGGTATATATTACTCTTCGTTTTTCATTTATCTTTTT
TTTTGTTGTTGATGGGAATGAAATTGCCTGGTCCTTTTGGTGTGTTTTTGAGCTTTTATT
ATTATACAGAGTGATCCCTTTTTTTGTTATAACTATTACAAGTTTTTAGCTTTATTTGATA
TGGATGCTCTCTCCTTTTTCTTCTATCTGTTTCTGGAAATTTTGACCTCATCATATTACTT
ATGTCATCCAAA

>G740 Amino Acid Sequence (domain in AA coordinates: 24-42, 232-268)

MASEDQSAARSTGKVNWFNASKGYFITPDDGSVELFVHQSSIVSEGYRSLTVGDAVEFA
ITQGSDBGKTKAVNVTPAGGSLKKENNSRNGARRGGSGCYNCGELGHISKDCGIGGG
GGGERRSRGEGCYNCGDTGHFARDCTSAGNGDQRGATKGGNDGCTCGDVGHVARDCT
QKSVNGDQRGAVKGGNDGCTCGDVGHFARDCTQKVAAGNVRSGGGSGTCYSCGGVGH
IARDCATKRQPSRGYQCGSGHLARDCDQRGSGGGGNDNACYKCKEGHFAECSSVA*
>G770 (119..1069)

CCTTCCCTATATAAGGAAGTTCATTTCAATTTGGAGAGGACACGCTGACAAGCTGACTCT
AGCAGATCTGGTACCGTCGACGGTTCTTGGATTGGAGTAAACTAAAGATCATATAAAAT
GGAACAAGGAGATCATCAGCAGCATAAGAAAGAAGAAGCTTTGCCACCGGGTTTCAG
ATTTTCATCCGACGGATGAGGAGCTAATCTCATATTACTTGGTTAATAAGATTGCCGATCA
AAACTTCACCGGAAAGCAATCGCTGACGTTGATCTTAACAAGTCCGAGCCATGGGAGCT
TCCTGAGAAGGCGAAAATGGGAGGAAAAGAATGGTACTTTTTTAGCCTCCGGGACCGGAA
GTACCCGACGGGAGTGAGGACGAATAGGGCGACGAATACAGGATATTGGAAAACACAGG
AAAAGACAAAGAGATATTCAATAGCACAACTCGGAGTTGGTTCGGGATGAAGAAGACTTT
GGTCTTTTACAGAGGACGAGCTCCTCGTGGGGAGAAGACTTGTGTTGGGTTCATGCATGAGTA
TCGACTTCACTCCAAGTCTCATATAGAACCTCCAAGCAAGACGAGTGGGTAGTGTGTAG
AGTGTTCAGAAAACAGAAGCAACCAAGAAATACATAAGCACCAGTAGCAGCAGCACAAG
TCATCACCACAACAACACACAAGAGCCTCAATACTATCAACCAACAACAATAATCTTAA
TTACTCATCAGACCTCCTTCAACTCCACCGCATCTACAACCACACCCGAGCCTCAATAT
TAACCAATCCCTCATGGCAAACGCCGTTACCTAGCTGAGCTCTCAAGAGTCTTCCGTGC
CTCTACAAGCACCACCATGGACTCTTCTCATCAGCAGCTAATGAAGTACACCCACATGCC
TGCTTCAGGGCTCAACCTCAACCTTGGCGGTGCAGTGGTCCAGCCGCTCCTGTTGTGTC
TCTTGAGGATGTTGCGCGGTTAGTGCTTCGTACAATGGCGAAAACGGGTTTGGAAATGT
GGAGATGAGCCAGTGCATGGACTTGGATGGATACTGGCCATCTTATTGATTGGTAATTGT
CAGTTTAAAGTTATGGTTTTTATATTGTTTCCATTTACTTGTGGTAAAACGATTTTGGTT
GTTCTTGCGAACGCTCTAGACAGGCCTCGTACCGGATCCTCTAGCTAGAGCTTTCGTTTCG
TATCATCGGTTTC

>G770 Amino Acid Sequence (domain in AA coordinates: 19-162)

MEQGDHQHKKKEEALPPGFRFHPTDEELISYYLVNKIADQNFTGKAIADVLDLNKSEPWE
LPEKAKMGGKEWYFFSLRDRKYPTGVRTNRTNTGYWKTGKDKKEIFNSTTSELVGMKKT
LVFYRGRAPRGEKTCWVMHEYRLHSKSSYRTSKQDEWVVCRVFKTEATKKYISTSSST
SHHHNNHTRASILSTNNNNPNYSSDLLQLPPLQPHPSLNINQSLMANAVHLAELSRVFR
ASTSTTMDSSHQQLMNYTHMPVSGNLNLGGALVQPPPVVSLDVDAAVSASYNGENGFN
VEMSQCMDLDGYWPSY*

>G858 (99..869)

CATAATCTCTTCTCTATATCTCTTCTTCTTCTTTTACCCTGTTTTTTTTTTCATTC
CACAGAGCCCAGGTTGATTGATTTTGTATTTCAGAGATATGGGGAGAGGAAGGATTGAGA
TTAAGAAGATTGAGAAATATCAACAGTCGTCAAGTCACCTTCTCTAAGAGACGAAACGGTT
TGATCAAGAAGGCTAAAGAGCTTTCGATTCTCTGTGACGCCGAGGTTGCTCTTATCATCT
TCTCCAGCACCGGCAAGATTTACGATTTCTCCAGCGTCTGTATGGAGCAAATTTCTTCTA
GATATGGATACACTACTGCGTCCACTGAGCATAAACAACAAGAGAACACCAACTTCTAA
TTTGTGCTTCACATGGAAATGAAGCTGTGTTGCGAAATGATGATTCTATGAAGGGGGAAC
TTGAAAGATTACAGCTTGCAATTGAGAGACTTAAGGGTAAGGAGCTTGAAGGTATGAGTT
TCCCGGATCTTATTTCTTTGAAAACAGTTGAACGAGAGCTTGCATAGTGTCAAGGATC
AAAAGACACAAATCTGTCTCAACCAGATTGAGAGATCCAGGATACAGGAGAAAAAGCAT
TGGAAGAAAACCAATCTTGCGCAAACAGGTTGAGATGTTGGGGAGAGGTTTCAGGACCAA
AAGTGTGTAATGAAAGGCCTCAAGATTCTAGCCCAGAAGCCGATCCCGAGAGCTTTCAT

CAGAAGAGGATGAGAATGACAACGAGGAGCACCATTCCGACACTTCCTTGCACTTGGGGT
TGTCGTCGACGGGGTATTGCACAAAGAGAAAGAGCCGAGATCGAACTGGTCTGCGATA
ACTCTGGGAGTCAAGTGGCTTCTGATTGATGGAATCGATTATTTTTCTAATTCTGGTTGT
TTAGGGGTCTCTATGTGTCTTCTGTTTCTGGCTGTTCTTTTGCTTTATTTTCATCTCAAG
TAGAGTTTTCTTAATGTTTAGGTGGAACATTTTCCATAATCAAGAAGGGATTGATCAA
TCAATAACATTAGATTTTCTTAGTTAAAGACTTAAAGTTGCCCCACACACCACCATATG
TGATTATGATGAATTTACATTTTATAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G858 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRGRIEIKKIENINSRQVTFSKRRNGLIKKAKELSILCDAEVALIIFSSTGKIYDFSSV
CMEQILSRGYTTASTEHHKQREHQLLICHASHGNEAVLRNDDSMKGELERLQLAIERLKG
KELEGMSFPDLISLENQLNESLHVKDQKTQILLNQIERSRIQEKKALEENQILRKQVEM
LGRGSGPKVLNERPDSSPEADPESSSSEEDENDNEEHSDTSLQLGLSSTGYCTKRKKP
KIELVCDNSGSQVAD*

>G865 (282..920)

ATCCCCACTTGTGTTCATCACCAGCCAGCTCCATGTCCTAGTCACTCCACAGATTCC
CTATCATCATCAATTCGTTTCAAACCTTAGTTTCCTTTCAAAGTCTTGACATATATACACA
CACACCTATTATTCTCTTGGTGTGTTTGTGTGTACATATACGTGTGAGTACATACTTTG
TTGTAAAGTGGATCGGAGGTATGGAAAGGACCGGTTCCACCGGAAACATCGGCGGCGG
CGGATGATAATTCGTCTTGAACGAGACTGATGTCACCGCCATGGTCTCCGCTCTCAGCC
GTGTCATAGAGAATCCGACAGACCCGCGGTCAAACAAGAGCTTGATAAATCGGATCAAC
ATCAACCAGACCAAGATCAACCAAGAAGAAGACACTATAGAGGCGTAAGGCAGAGACCAT
GGGGTAAATGGGCGGCAGAAATCCGCGATCCAAAGAAAGCAGCCCGTGTCTGGCTCGGGA
CTTTCGAGACGGCAGAGGAAGCTGCTTTAGCCTATGACCGAGCTGCCCTCAAATTCAAAG
GCACCAAGGCTAAACTGAACTTCCCTGAACGGGTCCAAGGCCCTACTACCACCAACCA
TTTCTCATGCACCAAGAGGAGTTAGTGAATCCATGAACTCACCTCCTCCTCGACCTGGTC
CACCTTCAACTACTACTACTTCTGCGGCAATGACTTATAACCAGGACATACTTCAATACG
CTCAGTTGCTTACGAGTAACAATGAGGTTGATTTATCATACTACACGTGACTCTCTTCA
GTCAACCTTTTTCAACGCCTTCTTCATCTTCTTCTCCCAACAGACGCAGCAACAGC
AGCTACAACAACAACAACAGCAGCGTGAAGAAGAAGAGAAGAAATTATGGTTACAATTATT
ATAACTACCCAAGAGAATAATCTAAT'TATTATTGTTGGTTCGAATCAGTTTTATAAATAGC
TATCATAGTTTTCATTTTGGTTTCCGTAACTTTGTTGCATGGAAAATATGAATGAACGA
GGGACATGTGTAACAATTGTTTGTGTTTTCGTAAATGTTAGTTGTATTGGATTGCTGA
AGTTTGATTTTCTGAGCATAAATCATTTGACGGTCAAAAAAAAAAAAAA

>G865 Amino Acid Sequence (domain in AA coordinates: 36-103)

MVSALSRVIENPTDPVVKQELDKSDQHQPDPDQPRRRHYRGVRQRPWGKWAIEIRDPKKA
ARVWLGTFFETAEEALAYDRAALKFKGTKAKLNFPERVQGPTTTTTISHAPRGVSESMNS
PPPRPGPPSTTTTWSWMTYNQDILQYAQLLTSNNEVDLSYYTSTLFSQPFSTPSSSSSSS
QQTQQQQQLQQQQQQREEEKKNYGYNYNYNPRE*

>G872 (59..646)

CCGAAACAGAAATCAACCAACCGAATCGAACCAGAACCGGAGTTTTTATCCAAT
GGTGAAGCAAGCGATGAAGGAAGAGGAGAAGAAGAGAAACACGGCGATGCAGTCAAAGTA
CAAAGGAGTGAGGAAGAGGAAATGGGGAAAATGGGTATCGGAGATCAGACTTCCACACAG
CAGAGAACGAATTTGGTTAGGCTCTTACGACACTCCCGAGAAGGCGGCGGTGCTTTCGA
CGCCGCTCAATTTTGTCTCCGCGGCGGCGATGCTAATTTCAATTTCCCTAATAATCCACC
GTCGATCTCCGTAGAAAAGTCGTTGACGCTCCCGAGATTCAAGGAAGCTGCTGCTAGATT
CGCTAACACAT'TCCAAGACATTGTCAAGGGAGAAGAAGAAATCGGGTTTTAGTACCCGGATC
CGAGATCCGACCAGAGTCTCCTTCTACATCTGCATCTGTTGCTACATCGACGGTGGATTA
TGATTTTTCGTTTTCGATTGCTTCCGATGAATTTCCGGTTTGATTCTTCTCCGACGA
CTTCTCTGGCTTCTCCGGTGGTGATCGATTACAGAGATTTTACCCATCGAAGATTACGG
AGGAGAGAGTTTATTAGATGAATCTTTGATTCTTTGGGATTTTGAATTCCCAAACATAA
TATTTTTTTAGAGCGAACTGTGAGATTTTCTTGGAGTCATGGAGAAATCTGGAGATTTT
TTGTAACACGGAGCTCCAATGACCCGGGAATTTCTTTTCGTTTCGGATCCGAATTTGATGT
GGATCATATTACACCTATATTTTTTTCATTTTTTTGTTGTAAGAAAAATCGGATAAGAT
TCTAGTAATAAATGTTAAAAAGTCCATTTTCAATTAAAAAAAAAAAAAAAAAAAAAA

>G872 Amino Acid Sequence (domain in AA coordinates: 18-85)

MVKQAMKEEEKKRNTAMQSKYKGVKRKWKWVSEIRLPHSRERIWLGSYDTPEKAARAF
DAAQFCLRGGDANFNFPNNPPSISVEKSLTPPEIQEAAARFANTFQDIVKGEESGLVPG

SEIRPESPSTSASVATSTVDYDFSFLDLLPMNFGFDSFSDDFSGFSGGDRFTEILPIEDY
GGESLLDESILWDF*

>G904 (1..1005)

atggaatctctcatcaatcccagccatggcggaggaaactacgattctcactcttcttct
ctcgatagtcctcaaaccaagcgtactagtcattctctcattctcctcatgactcttctc
atctccggtttccatttgccttctcctccgctgtctcaatcgctgtagccaccgctccggt
ctccctctttcatcttcttctcctccgctcgcaaccgtaacttccgattcccagcattctct
ggacatcgagtcctctcccgaacagaacgggtcctccgctgcttgattcgcttccgattttc
aaattctcctccgctcactcgccgatctagctccatgaattccggagattgcccgtttgt
ttgtcgaaattcgaaaccggaggatcagctccgtcttcttctcctctctgttgacgctttt
cacgccgattgtatcgatatctggctagctcttaaccagacttgctctctctgctgctct
cctctcttctgcttcagaatctgactctcatgaagtctctcgccgtcgctcggtcaaacaac
ggcggaggagaaaacagcttccgtctcgaaatcggtatccatcagccgtcgctcgtaacaac
ccgattccagaatccggttagcagcagcgaacttactcaatcggttcggttcgattacata
gtagacgagctagattcagaaatctcagagtcgaatttcaaccgtggaaaacaggaagac
gcgactacaacaactgccacagcaacggcggttacgactaatccgacgtcggttgagct
agtttagcggcgatataaggaacgatgggtctagaagctggctcaaggattacgttgac
agactctcacgaggtatctcgtcgctgcaatgtcggttagaagctctggtagattttt
actgggagtagtcgctggagtgaggaattgacggtgatggatttagaagcgaatcatgcc
ggagaagagataaagtgaagctttccggtggctctcaggggtgtga

>G904 Amino Acid Sequence (domain in AA coordinates: 117-158)
MESLINPSHGGGNYDSHSSSLDSLKPSVLVIIILILLMTLLISVSICFLRLNRCSHRSV
LPLSSSSSVATVTSDSRFSGHRVSPETERSSVLSLPIFKSSVTRRSSMNSGDCAVC
LSKFEPEDQLRLPLCCHAFHADCIDIVLSNQTCPLCRSPLFASESILMKSLAVVGSNN
GGGENSFRLIGSISRRRQTPIPESVEQHRTYSIGSFYIVDDVDSEISESNFNRGKQED
ATTTTATATAVTTNPTSFEASLAADIGNDGSRSLKDYVDRLSRGISSRAMSFRSSGRFF
TGSSRRSEELTVMLEANHAGEEISELFRWLSGV*

>G910 (1..1071)

ATGTTATGTATAATAATAATTGAGAATATGGAAGAGTATGTGAGTTTGTAAAGCGTAT
AGAGCAGTGTTTTATTGTATAGCTGATACAGCAAATCTTTGTTTAAATGTGATGCAAAG
GTTCAATTCAGCTAATCTACTCTCGGGACGGCATTACGTACGGTTTATGTGATTCTGGT
AAGAATCAGCCTTGTGTGTCGGATGTTTGGACCATAAAATGTTCTTTGCCATGGATGT
AATGATAAGTTTCATGGTGGTGGCTCTCTGAGCATCGTAGAAGGGATTGAGGTGTTAT
ACGGGTTGTCTCTCTGCTAAAGATTTCGCGGTTATGTGGGGTTTTCGAGTTATGGATGAC
GATGATGATGTTTTCGTTAGAGCAATCTTTTGAATGGTTAAACCTAAGGTGCAAAGAGAA
GGTGGTTTTATCTTGAACAGATTCTTGAATTGGAGAAGGTTTCAGCTCAGGGAAGAGAA
GGTAGTTCTTCTTGAACAGAGGATCCATCTCCATTGGAGCTTCTTAAGAAACCC
GAAGAACAGTTAATCGATCTTCCGACAGACCGGAAAGAGCTGGTTGTTGATTTTTCACAC
TTGTCTCATCTTCCACACTTGGTGATTCTTTTGGGAATGCAAAAGTCCATACAATAAG
ACAATCAGTTGTGGCATCAAAATATACAAGACATTGGAGTATGTGAAGATACAATCTGC
AGTGACGATGACTTCCAATACCTGACATTGATCTCACTTTCCGGAACTTTGAAGAGCAA
TTTGGAGCTGATCCTGAGCCAATTGCAGATAGTAACAACGTGTTCTTTGTTTCTTCCCTT
GACAAATCAGATGAGATGAAGACATTTTCTTCTTCAATAATCCCATATTTGCACCT
AAACCAGCTTCACTCAATATCTCATTCTCAAGCAGTGAAACCGATAACCCTTATAGTCAC
TCAGAGGAAGTAATCTCATTTTGTCCCTCCCTCTTAACAATACACGTCAAAGGTCATC
ACAAGGCTCAAGGAGAAGAAGAGAGCAAGAGTGGAGGAGAAAAAGCTTAA

>G910 Amino Acid Sequence (domain in AA coordinates: 14-37, 77-103)
MLCIIIIENMERVCFCKAYRAVVYCIADTANLCLTCDKAVHSANSLSGRHLRTVLCDSG
KNQPCVVRCFDHKMFCHGKNDKFHGGSSSEHRRDLRCYTGCPKDFAVMWGFRVMD
DDVSLQSFVMPKPVQREGGFLEQILELEKVQLREENGSSSLTERGDPSPLEPKP
EEQLIDLPTGKELVDFSHLSSSSTLGDSEWCKSPYNKNNQLWHQNIQDIGVCEDTIC
SDDDFQIPDIDLTFRNFEQFGADPEPIADSNVFFVSSLDKSHMKTFSSSFNNPIFAP
KPASSTISFSSSETDNPYSHSEVISFCPSLSNNTRQKVITRLKEKKRARVEEKK*
>G912 (20..694)

CATCTTATCCAAAGAAAAATGAATCCATTTTACTCTACATTCCCAGACTCGTTTCTCTC
AATCTCCGATCATAGATCTCCGGTTTCAGACAGTAGTGAGTTTACCAAAGTTAGCTTC
AAGTTGTCCAAAGAAACGAGCTGGGAGGAAGAAGTTTCGTGAGACAGTCATCCGATTTA

CAGAGGAGTTCGTCAGAGGAATCTCTGGTAAATGGGTTTGTGAAGTTAGAGAGCCTAATAA
GAAATCTAGGATTTGGTTAGGTACTTTTCCGACGGTTGAAATGGCTGCTCGTGCTCATGA
TGTTGCTGCTTTAGCTCTTCGTGGTTCGCTCTGCTTGTCTCAATTCGCTGATTCTGCTTG
GCGGCTTCGTATTCTGAGACTACTTGTCTTAAGGAGATTACAGAAAGCTGCGTCTGAAGC
TGCAATGGCGTTTCAGAAATGAGACTACGACGGAGGGATCTAAACTGCGGCGGAGGCAGA
GGAGGCGGCAGGGGAGGGGGTGAAGGAGGGGAGAGGAGGGGCGGAGGAGCAGAATGGTGG
TGTGTTTTATATGGATGATGAGGCGCTTTTGGGGATGCCCAACTTTTTTTGAGAATATGGC
GGAGGGGATGCTTTTGGCGCCGCCGGAAGTTGGCTGGAATCATAACGACTTTGACGGAGT
GGGTGACGTGTCACTCTGGAGTTTACGAGTAATTTTTTGGCTCTTTTTCTGGATAATA
AGTT

>G912 Amino Acid Sequence (domain in AA coordinates:51-118)
MNPFYSTFPDSFSLISDHRSPVSDSSECSPKLASSCPKRAGRKKFRETRHPIYRGVRQR
NSGKWVCEVREPNNKSRIRWLGTFTVEMAARAHDAALALRGRSACLNFAWSRLRIPE
TTCPEIKQAASEAAMAFQNETTEGSKTAAEAEEAAGEGVREGERRAEQNGGVFYMD
EALLGMPNPFENMAEGMLLPPEVGVNHNDFDGVGDVSLWSFDE*

>G920 (114..1154)
AAAAATCTATTTTCTTCTCTTCCACTATATTACAACATTTCTTCATTCTCAAATCATC
ATACTAAAAACCTAAAAAAGTTACATATTCATTGTATCTTTGTGAGAAAAAATGGATT
CGAATAGTAACAACACGAAATCCATAAAGAGAAAAAGTTGTGACCAACTTGTGGAAGGCT
ATGAATTCGCTACTCAGCTTCAGCTTCTCCTTTCTCATCAACACTCTAACCAGTACCACA
TCGATGAGACCCGTCTTGTTTCCGGGTCGGGTTTCAGTTTCCGGTGGTCCAGATCCCGTTG
ATGAGCTCATGTCTAAGATCTTGGGATCTTTCCATAAACTATATCGGTTCTTGATTCTT
TTGATCCCGTCGCCGTCTCTGTCCCATCGCCGTCGAGGGTTTCATGGAATGCTTCATGTG
GGGATGATTCGGCGACTCCGGTGAGTTGCAACGGTGGAGATTCCGGTGAGAGTAAGAAGA
AGAGATTAGGGGTTGGTAAGGGTAAAGAGGATGCTACACTAGAAAGACGAGATCACATA
CAAGGATCGTGAAGCTAAAGTTCTGAAGACAGATATGCTTGGAGGAAATATGGACAAA
AGGAGATTCTTAATACCACATTCCCAAGAAGTTACTTTAGATGCACACACAAGCCAACGC
AAGGATGCAAAGCAACAAAGCAAGTTCAGAAACAGGATCAAGATTCTGAGATGTTCCAAA
TCACATACATTGGCTACCACACATGCACTGCCAATGACCAAACGCACGCGAAGACCGAGC
CTTTTGATCAAGAAATCATTATGGATTCCGAAAAGACATTGGCTGCTAGCACTGCTCAGA
ACCATGTCAATGCTATGGTGCAAGAGCAAGAGAACAACACCAGCAGTGTGACAGCAATAG
ACGCAGGCATGGTTAAGGAGGAACAAAATAACAATGGTGATCAGAGTAAAGATTATTATG
AGGGCTCTTCGACAGGTGAGGACTTGTCTATTGGTTTGGCAAGAGACGATGATGTTTGATG
ATCATCAAAATCACTACTATTGTGGTGAAACCAGTACTACTTCTCATCAATTTGGTTTCA
TCGACAACGATGATCAGTTTTCTCCTTCTTCGACTCATATTGTGCTGATTATGAAAGAA
CAAGTGCTATGTGAACATCCAAATCTGGAATGATGAATCAGCACTAGGTCTTCTCTTGA
GTATGTCTAGTTTAAATGTAATATTTTGTGTATGTTTGATAAAAAACACCATATATACTT
CTCTTTTACACCAAAAAAAAAAAAAAAAAAAAAA

>G920 Amino Acid Sequence (domain in AA coordinates: 152-211)
MDSNSNNTKSIKRKVVQDLVEGYEFATQLQLLLSHQHSNQYHIDETRLVSGSGSVSGGPD
PVDELMSKILGSFHKTIISVLDSDPVAVSVPPIAVEGSWNASCGDSDSATPVSCNGGDSGES
KKKRLGVGKGRGCTYTRKTRSHRIVEAKSSEDRYAWRKYQKEILNTTFPRSYFRCTHK
PTQGCKATKQVQKQDQDSEMFQITYIGYHTCTANDQTHAKTEPFDQEIIMDSEKTLAAS
AQNHVNAMVQEQENNTSSVTAIDAGMVKEEQNNNGDQSKDYEGSSTGEDLSLVWQETMM
FDDHQNHYYCGETSTTSHQFGFIDNDDQFSSFFDSYCADYERTSAM*

>G939 (9..1565)
CAGATTCTATGGATATGTATAACAACAATATAGGGATGTTCCGGAGTTTAGTTTGTAGCT
CGGCGCCTCCATTTACAGAGGGACATATGTGTTCTGATTTCGCATACGGCTTTGTGCGATG
ATCTGAGTAGTATGATGAGGAAATGGAATAGAGGAGCTTGAGAAGAAGATCTGGAGAGACA
AGCAGCGTTTAAAGCGGCTCAAGGAAATGGCGAAGAACGGTCTAGGAACAAGATTGTTGT
TGAAGCAGCAACATGATGATTTTCCAGAGCACTCTAGTAAGAGAACCATGTACAAGGCAC
AAGATGGGATCTTGAAGTACATGTGCAAGACAATGGAGCGATATAAAGCTCAAGGTTTGT
TTTATGGGATTGTGTAGAGAATGGGAAACGGTAGCGGGATCTTCTGATAATCTCCGTG
AATGGTGGAAGACAAAGTGAGGTTTGTATAGGAACGGCCAGCTGCTATAATCAAGCACC
AAAGGGATATCAATCTTCTGATGGAAGTGATTACAGGCTCTGAGGTTGGGGATTCTACCG
CACAGAAGTTGCTTGAGCTTCAAGATACTACTCTTGAGCTCTGTTATCGGCTCTGTTTC
CTCACTGCAACCCCTCTCAGAGGCGGTTTCCGTTGGAGAAAGCGTGACACCGCCATGGT

GGCCAACGGGGAAAGAAGATTGGTGGGATCAACTGTCTTTACCCGTTGATTTTCGAGGTG
TTCCGCCACCTTACAAGAAGCCTCATGATCTCAAGAAGCTGTGGAAAATTGGTGTTTGA
TTGGTGTAAATCAGACATATGGCTTCTGACATTAGCAACATACCCAATCTCGTGAGACGGT
CTAGAAGTTTGCAGGAGAAAATGACGTCAAGAGAAGGCGCTTTATGGCTCGCTGCTCTTT
ACCGAGAAAAGGCTATTGTTGATCAAATAGCCATGTCTAGAGAAAACAACAACACTTCTA
ACTTTCTTGTTCCTGCAACCGGTGGAGACCCAGATGTTTTGTTTCTGAATCTACAGACT
ATGATGTTGAACTGATTGGTGGCACTCATCGGACCAATCAGCAGTATCCTGAATTTGAAA
ACAATACTAACTGTGTTTACAAGAGAAAGTTTGAAGAAGATTTTGGGATGCCAATGCATC
CAACACTCCTAACATGTGAGAACAGTCTCTGTCTTATAGCCAACCATATGGGATTTCT
TTGACAGGAACTTAAGAGAGAATCACCAAATGACTTGTCTTATAAAGTCACTTCCTTCT
ACCAACCAACTAAACCCTATGGTATGACGGGTTTAATGGTTCCTTGTCCGGATTATAACG
GGATGCAGCAGCAGGTTTCAAGACCAGTTAATCATCCCAACGATCTCTACA
GACCAAAGCTCCACAAGAGGCAACGATGACTTGGTTGAGGATTTGAATCCTTCTCCTT
CGACGCTGAATCAGAATCTTGGTTAGTCTTACCTACTGACTTCAATGGAGGTGAGGAAA
CAGTAGGAACAGAGAACAACTCTGCATAATCAAGGGCAAGAGTTGCCACATCTTGGATT
AGTAAAGAAAGCTTCAAGTCTTCTTTTATGTTTTCTAGTCTTTATAGCTTTGTCTCTT
GCTTATTCTCTCATTAAACACAGTCTTCTGATCTCTCCATTTATAGCCCATGTAGCAATG
GAGAAAGATTAGGTTTTCATAATAAGTTAATAACCAAATTCAAA

>G939 Amino Acid Sequence (domain in AA coordinates: 97-106)
MDMYNNNIGMFRSLVCSAPPFTEGHMCSDSHTALCDDLSSDEEMEIEELEKKIWRDKQR
LKRLEMAKNGLGTRLLLLKQHHDFPEHSSKRTMYKAQDGLKYSKTMERYKAQGFVYV
IVLENGKTVAGSSDNLRFHWKDKVFRDRNGPAIIKHQDINLSDGSDSGSEVGDSTAQK
LLELQDPTLGLALLSALFPHCNPPQRRFPLEKGVTPPWPTGKEDWWDQLSLPVDVFRGVPP
PYKKPHDLKKLWKIGVLIGVIRHMASDISNIPNLVRRSRSLQEKMTSREGALWLAALYRE
KAIVDQIAMSRENNNTSNFLVPATGGDPDVLFPSTDYDVELIGGTHRTNQQYPEFENNY
NCVYKRKFEEFDGMPMHPHTLLTCENSLCPYSQPHMGFLDRNLRENHQMTCPIKVTSTFYQP
TKPYGMTGLMVPDPCPDYNGMQQVQSFQDQFNHPNDLYRPKAPQGRNDDLVEDLNPSSTL
NQNLGLVLPDFTNGGEETVGTENNHLNQGOELPTSWIQ*

>G963 (1..897)

ATGAGTTTGCCTCCAGGATTCAAGTTTCATCCCACTGATGAAGAAGCTGGTGGCTTACTAT
CTTGATAGGAAGGTCAACGGCCAAGCCATTGAGCTCGAGATCATCCGAGAAGTTGATCTT
TATAAATGCGAGCCATGGGACTTGCCTGAAAAGTCAATTTTGGCCGGGAAACGACATGGAA
TGGTACTTTTACAGCACAAAGGGATAAGAAGTATCCAAATGGCTCTAGGACGAACCGTGCG
ACCCGAGCGGGTTACTGGAAGGCCACGGGGAAGATCGTACAGTAGAATCAAAGAAGATG
AAGATGGGAATGAAGAAGACACTGGTTTATTATAGAGGAAGGGCTCCTCATGGCCTTCGT
ACTAATTGGGTCTATGCATGAATATCGTCTCACGCACGCTCCTTCTCTCTCTTGAAGGAG
TCGTATGCATTGTGCCGAGTGTTTAAGAAGAACATACAAATTCCAAAGAGAAAAGGGGAA
GAAGAAGAAGCAGAAGAAGAGAGCACTAGTGTAGGAAAAGAAGAGGAAGAAGAAAAGGAG
AAGAAGTGGAGAAAATGTGATGGTAATTATATTGAAGACGAGAGCTTGAAAAGAGCATCC
GCGGAGACATCTTCATCAGAGCTAACTCAAGGGTCTTTTATAGACGAAGCAAACAGCTCA
TCCATATTTGCTCTTCATTTCTCATCTTCTCTTCTGGACGATCATGATCATCTTTCTCA
AACTATTCTCATCAGCTTCCATATCATCTCTCTTCAACTCCAAGATTTCCCTCAACTT
TCTATGAACGAAGCAGAGATTATGTCAATCCAACAAGACTTTCAATGCAGAGACTCTATG
AACGGGACACTTGACGAAATCTTCTCTTCTTCCGCCACTTTCCCGCTTCCCTTTGA

>G963 Amino Acid Sequence (domain in AA coordinates: TBD)
MSLPPGFRFHPTDEELVAYYLDKRVNGQAIIELEIPEVDLYKCEPWLPEKSFPLPGNDME
WYFYSTRDKKYPNGSRTNRATRAGYWKATGKDRTVESKKMKMGKTLVYYRGRAPHGLR
TNWVMHEYRLTHAPSSSLKESYALCRVFKNIQIPKRKGESEEBEESTSVGKEEEEEKE
KKWRKCDGNYIEDESLKRASAEISSSELTQGVLLDEANSSSIFALHFSSSLDDHDLHLS
NYSHQLPYHPLQLQDFPQLSMNEAEIMSIQQDFQCRDSMNGTLDEIFSSSATFPASL*

>G979 (60..1352)

CCTCTGAGGAATCAAATCACTCACACTCCAAAAAAATCTAAACTTTCTCAGAGTTTAA
TGAAGAAGCGCTTAACCACTTCCACTTGTCTTCTTCTCCATCTTCCTCTGTTTCTTCTT
CTACTACTACTTCTCTCTTATTCAGTCGGAGGCTCCAAGGCCTAAACGAGCCAAAAGGG
CTAAGAAATCTTCTCTTCTGGTGATAAATCTCATAACCCGACAAGCCCTGCTTCTACCC
GACGACGCTTATCTACAGAGGAGTCACTAGACATAGATGGACTGGGAGATTGAGGCTC
ATCTTTGGGACAAAAGCTCTTGGAAATTCGATTCAGAACAAGAGGCAACAAGTTTATC

AGAGATACGGTTTTTGAAGCTGATAAGGAAGATAAAGCCAGACGTGTTTCATCCCCGGGATC
 CTCAGCGGATCCTACAACGCGCCTTTCTTTGTCACGAGGTTTAGAGAAGTTCTGTTTCAT
 TACTCATCTCTGTTTGACATGTGTGACACGAATCTAACACGGGAAGATCCAATGAGGGTT
 ATGTTTGAGAAAGAGTTCTATGGGCGGGAGATCATGAACGTGGTGGCGTGTGAGGGGACG
 GAGAGAGTGGAGAGGCCAGAGAGTTATAAGCAGTGGCAGGCGAGGGCGATGAGAGCCGGG
 TTTAGACAGATTCCGCTGGAGAAGGAAGTAGTTTCAAGAACTGAAGTTGATGGTGGAAAGT
 GGATACAAACCCAAAGAGTTTGATGTTGATCAAGATTGTCACCTGGTTGCTTCAGGGCTGG
 AAAGGTAGAATTGTATACGGTTTCATCTATTTGGGTTCTTTCTTTTCTATGTGGGCAGA
 GCAACTAGGGTTTTGATCATGGATCCAAACTTCTCTGAATCTCTAAACGGCTTTGAGTAT
 TTTGATGGTAACCCCTAATTTGCTTACTGATCCAATGGAAGATCAGTATCCACCACCATCT
 GATACTCTGTTGAAATACGTGAGTGAGATTCTTATGGAAGAGAGTAATGGAGATTATAAG
 CAATCTATGTTCTATGATTCATTGGCTTTACGAAAACTGAAGAAATGTTGCAGCAAGTC
 ATTACTGATTCTCAAATCAGTCCTTTAGTCCTGCTGATTTCATTGATTACTAATTCTTGG
 GATGCAAGCGGAAGCATCGATGAATCGGCTTATTCGGCTGATCCGCAACCTGTGAATGAA
 ATTATGGTTAAGAGTATGTTTAGTGATGCAGAATCAGCTTTACAGTTTAAAGAAAGGGGT
 GAAGAAGCTAGTAAATCCTTCCCAATAGTGATCAATGGGTATCAATCTGGATATCGAG
 AGATCCGAAAGGCGGATTCGGTTAAAGAAGAGATGGGATGGATCAGTTGAGAGTTAAG
 AAGAATCATGAAAGGGATTTTGAGGAAGTTAGGAGTAGTAAGCAATTTGCTAGTAATGTA
 GAAGATAGTAAGGTTACAGATATGTTTGATAAGGTTTTCCTTCTTGACGGTGAATGCGAT
 CCGCAACATTGTTAGACAGCGAGATTCAAGCGATTCCGAGTAGTAAGAACATAGGAGAG
 AAAGGGAAGAAGAAGAAGAAGAAGAGTCAAGTGGTTGATTTTCGTACACTTCTCACT
 CATTGTGCACAAGCCATTTCCACAGGAGATAAAACCACGGCTCTTGAGTTTCTGTTACAG
 ATAAGGCAACAGTCTTCGCCTCTCCGTGACGCGGGGCAAAGACTAGCTCATTTGTTTCGCT
 AACGCGCTTGAAGCTCGTCTACAGGGAAGTACCGGTCCTATGATCCAGACTTATTACAAT
 GCTTTAACCTCGTCTGTTGAAGGATACTGCTGCGGATACAATTAGAGCGTATCGAGTTTAT
 CTTTCTTCGCTCCGCTTTGTTACCTTGATGTATTTCTTCTCCATCTGGATGATTCTTGAT
 GTGGCTAAAGATGCTCCTGTTCTTCATATAGTTGATTTTGGGATTCTATACGGGTTTCAA
 TGGCCGATGTTTATTAGTCTATATCAGATCGAAAAGATGTACCGCGGAAGCTGCGGATT
 ACTGGTATCGAGCTTCTCAGTGCGGGTTTCGGCCCCGCGAGCGAATAGAGGAGACAGGA
 CGGAGATTGGCTGAGTATTGTAAACGGTTTAAATGTTCCGTTTGAGTACAAAGCCATTGCG
 TCTCAGAACTGGGAAACAATCCGGATAGAAGATCTCGATATACGACCAAACGAAGTCTTA
 GCGGTTAATGCTGGACTTAGACTCAAGAACCTTCAAGATGAAAACAGGAAGCGAAGAGAAT
 TGCCCGAGAGATGCTGTCTTGAAGCTAATAAGAAACATGAACCCGGACGTTTTCATCCAC
 GCGATTGCTCAACGGTTCATTCAACGCACCCCTTCTTTATCTCGCGGTTTAAAGAAGCGGT
 TACCATTACTCCGCTCTCTTCGACATGTTTGATTTCGACGTTGCCTCGGGATAACAAAGAG
 AGGATTAGGTTCCGAGAGGGAGTTTACGGGAGAGAGGCTATGAACGTGATAGCGTGCGAG
 GAAGCTGATCGAGTGGAGAGGCCCTGAGACTTACAGGCAATGGCAGGTTAGAATGGTTAGA
 GCCGGGTTTAAAGCAGAAAACGATTAAAGCCTGAGCTGGTAGAGTTGTTTAGAGGAAAGCTG
 AAGAAATGGCGTTTACCATAAAGACTTTGTGGTTGATGAAAATAGTAAATGGTTGTTACAA
 GGCTGGAAGGTCGAACCTCTCTATGCTTCTTCTTGTGGGTTCTCTGCCTAG
 >G987 Amino Acid Sequence (domain in AA coordinates: 428-432, 704-708)
 MGSYSAGFPGLDWFDFPGLNGSYLNDQPLLDIGSVPPPLDPYPQONLASADADFSDSV
 LKYSISQVLMEEDMEDKPCMFHDALSQAAEKSLYEALGEKYPVDDSDQPLTTTTSLAQLV
 SSPGSSSYASSTTTSSDSQWSFDCLENNRPSSWLQTPIPSNFIFQSTSTRASSGNAVFG
 SSFSGDLVSNMFNDTDLALQFKKGMEASKFLPKSSQLVIDNSVPNRLTGKKSHWREEH
 LTEERSKKQSAIYVDETDELTDMDNLIIFGEAKEQPVCILNESFPKEPAKASTFSKSPK
 GEKPEASGNSYTKETPDLRMTLVSCAQAVSINDRRTADELLSRIQHSSSYGDGTERLAH
 YFANSLEARLAGIGTQVYTALSSKKTSTSDMLKAYQTYISVCPFKKIAIIFANHSIMRLA
 SSANAKTIHIIDFGISDGFQWPSLIHRLAWRRGSSCKLRITGIELPQRGFRPAEGVIETG
 RRLAKYQCQFNIPFEYNAIAQKWESIKLEDLKLKEGEFVAVNSLFRFRNLLDETVAHVSP
 RDTVLKLRIRKIPDVFIPIGLSGSYNAPFFVTRFREVLPHYSSLFDMCDTNLTREDPMRV
 MFEKEFYGREIMNVVACEGTERVERPESYKQWQARAMRAGFRQIPLEKELVQKLKLMVES
 GYKPKFEFDVDQDCHWLLQGWKGRIVYGSSIWVPPFFYVGRATRVLIMDPNFSESNGFEY
 FDGNPNLLTDPMEDQYPPPSDTLLKYVSEILMEESNGDYKQSMFYDSLALRKTEMLQQV
 ITDSQNQSFSFSPADSLITNSWDASGSIDESAYSADPQPVNEIMVKSMFSDAESALQFKKV
 EEASKFLPNSDQWVINLDIERSERRDSVKEEMGLDQLRVKKNHERDFEEVRSSKQFASNV
 EDSKVTDMFDKVVLLLDGECDPQTLDDSEIQAIRSSKNIGEKGKKKKKKKSQVVDFTLLT

HCAQAISTGDKTTALEFLLQIRQSSPLGDAGQRLAHCFANALEARLQGSTGPMIQTYYN
ALTSSLKDTAADTIRAYRVYLSSSPFVTLMYFFSIWMILDVAKDAPVLHIVDFGILYGFQ
WPMFIQSISDRKDVPRKLRLITGIELPQCGFRPAERIEETGRRLAEYCKRFNVPFEYKAIA
SQNWETIRIEDLDIRPNEVLAVNAGLRLKLNQDETGSEENCPRDAVLKLRNMNPDVFIH
AIVNGSFNAPFFISRFKEAVYHYSALFDMFDSTLPRDNKERIRFEREFYGREAMNVIACE
EADRVERPETYRQWQVRMVRAGFKQKTIKPELVELFRGKLKKWRYHKDFVVDENSKWLLQ
GWKGRTLYASSCWVPA*

>G993 (6..1091)

CAAATATGGAATACAGCTGTGTAGACGACAGTAGTACAACGTCAGAATCTCTCTCCATCT
CTACTACTCCAAAGCCGACAACGACGACGAGAGAAGAACTCTCTCTCCGCCGCGACGT
CGATGCGTCTCTACAGAATGGGAAGCGGCGGAAGCAGCGTCGTTTGGATTACAGAGAACG
GCGTCGAGACCGAGTCACGTAAGCTTCCTTCGTGCGAAATATAAAGCGGTTGTGCCTCAGC
CTAACGGAAGATGGGGAGCTCAGATTTACGAGAAGCATCAGCGAGTTTGGCTCGGTACTT
TCAACGAGGAAGAAGAAGCTGCGTCTTCTTACGACATCGCCGTGAGGAGATTCCGCGGCC
GCGACGCCGTCATACTTCAAATCTCAAGTTGATGGAACGACGCCGAATCGGCTTTTC
TTGACGCTCATTCTAAAGCTGAGATCGTGGATATGTTGAGGAAACACACTTACGCCGATG
AGTTTGAGCAGAGTAGACGGAAGTTTGTAAACGGCGACGGAACGCTCTGGGTTGGAGA
CGGCGACGTACGGAACGACGCTGTTTGTAGAGCGCGTGAGGTTTGTTCGAGAAGACTG
TTACGCCGAGCGACGTGCGGAAGCTGAACCGTTTAGTGATACCGAAACAACACGCGGAGA
AGCATTTCGCTTACCGCGATGACGACGCGGATGGGGATGAATCCGTCTCCGACGAAAG
GCGTTTGTATTAACTTGAAGATAGAACAGGGAAGTGTGGCGGTTCCGTTACAGTTACT
GGAACAGCAGTCAAAGTTACGTGTTGACCAAGGGCTGGAGCCGTTTCGTTAAAGAGAAGA
ATCTTCGAGCCGGTGATGTGGTTTGTTCGAGAGATCAACCGGACCAGACCGGCAATTGT
ATATCCACTGGAAGTCCGGTCTAGTCCGGTTCAGACTGTGGTTAGGCTATTCGGAGTCA
ACATTTTCAATGTGAGTAACGAGAAACCAAACGACGTCGAGTAGAGTGTGTTGGCAAGA
AGAGATCTCGGGAAGATGATTTGTTTTTCGTTAGGGTGTTCCAAGAAGCAGGCGATTATCA
ACATCTGTGACAAATCTTTTTTTTTTGGTTTTTTCTTCAATTGTTTCTCCTTTTTTCA
ATATTTTGTATTGAAATGACAAGTTGTAAATTAGGACAAGACAAGAAAAAATGACAACTA
GACAAAATAGTTTTTGTAAAAA

>G993 Amino Acid Sequence (domain in AA coordinates: 69-134)

MEYSCVDDSTTSSELSISTTPKPTTTEKKLSSPPATSMRLYRMGSGSSSVLDSSENGV
ETESRKLPSKSKYKGVVPQPNRWGAQIYEKHQRVWLGTFFNEEEEAASSYDIAVRRFRGRD
AVTNFKSQVDGNDAESAFDLAHSKAEIVMLRKHTYADEFEQSRKRVNGDGKRSGLETA
TYGNDAVLRAREVLFEKTVTPSDVGKLNRLVIPQHAKEKHFPLPAMTTAMGMNPSPTKGV
LINLEDRTGKVVFRFRYSYWNSSQSYVLTGWSRFVKEKNLRAGDVVCFERSTGPDRLQYI
HWKVRSSPVQTVVRLFGVNI FNVSNKPNDAVECVGKKRSREDDLFSLGCSKKQAIINI
L*

>G681 (1..804)

ATGGGGAGGACGACATGGTTCGACGTCGACGGGATGAAGAAAGGAGAGTGGACGGCAGAG
GAAGACCAGAAGCTCGGCGCTTACATCAACGAGCATGGCGTTTGTGATTGGCGTTCCCTC
CCCAAAGAGCTGGTTTGCAGAGATGTGGAAAGAGCTGCAGATTAAGGTGGCTTAACATAT
CTAAAGCCTGGGATTAGAAGAGGCAAATTCCTCCTCAAGAAGAAGAAGAAATCATCCAA
CTTCATGCTGTTCTCGGAAACAGGTGGGCGAGCCATGGCGAAGAAGATGCAGAATCGAACA
GACAATGATATCAAGAACCATTGGAACCTTGTCTCAAGAAAAGACTTTCGAGAAAGGGA
ATCGACCCTATGACCCACGAGCCCATCATCAACACCTCACCGTCAATACCACTAACGCA
GATTGTGGTAACTCTTCCACCACGACGTCCCGTCGACGACGGAAGCTCTCCTTCTCCTC
GGCTCGTCTCGTCTTCTTAACAAACTCGCCGCGAGGTATCTCATCTAGACAACATAGTCTC
GATAGGATCAAGTAATCTTGTCTGAATTCATAATCGAAAGCAGTGATCAAGCAAAAGAG
GAAGAAGAAAAAGAAGAAGAAGAAGAAGAGATTCAATGATGGGTGAGAAGATTGAC
GGTAGTGAAGGAGAAGATATTCAGATTTGGGGCGAGGAGGAAGTTAGGCGTTTAAATGGAG
ATTGATGCAATGGATATGTACGAGATGACTTCGTACGACGCTGTCATGTACGAGAGTAGT
CACATACCTGATCATCTCTTTTGACTTAATATAGTGTGACTGTGTGAGTGCATGCATGTT
>G681 Amino Acid Sequence (domain in AA coordinates: 14-120)

MGRTTWFDVDMKKGEWTAEEDQKLGAYINEHGVCWRSRSLPKRAGLQRCGKSCRLRWLNY
LKPGIRRGKFTPEEEEEIIQLHAVLGNRWAAAMAKKMQNRDNDIKNHNWSCCLKRLSRKG
IDPMTHEPIIKHLTVNTTNADCGNSSSTTSPSTTESPSSSGSSRLLNKLAAGISSRQHS
DRIKYILSNSIISSDQAKEEEEEKEEEEEERDSMMGQKIDGSEGEDIQIWGEEVEVRLME

IDAMDMYEMTSYDAVMYESSHILDHLF*

>G1482 (1..996)

ATGAAGATCAGGTGCGACGCTGCGATAAAGAAGAAGCGTCCGGTGTTCACGCGCCGAC
GAAGCATCTCTGCGGCGGCTGCGACCACCAAGTCCACCACGCTAACAAACTCGCCTCT
AAACATCTCCGTTTCTCTCTCCTTTATCCTTCTTCTTCCAACACCTCCTCTCTCTGCG
GACATCTGTGAGGATAAAAAAGCTCTGTTGTTCTGTCAACAAGATAGAGCTATTTTATGC
AAAGATTGCGATTTCATCGATCCACGCTGCGAACGAACACACAAAGAAACACGATAGGTTT
CTTCTTACAGGGGTTAAGCTCTCTGCAACATCGTCTGTTTACAAACCTACTTCGAAATCT
TCTTCTTCTTCTTCAAGCAACCAAGATTTCTCTGTCCCTGGATCATCAATCTCTAATCCT
CCTCCTCTCAAGAAACCTCTCTCAGCTCCTCCTCAGAGCAACAAGATCCAACCTTTTCG
AAGATCAACGGCGGTGATGCGTCCGTGAATCAGTGGGGATCCACAAGCAGATTCTCTGAG
TATTGATGGATACGTTACCTGGTTGGCAGTTGAGGATTTCTCTGATTCCTCTCTTCTCT
ACTTATGGTTTCTCTAAGAGTGGTGATGATGATGGAGTGTACCATATATGGAACCAGAA
GATGACAACAACACTAAGAGAAACAACAACAACAACAACAACAACAACAATACAGTG
TCACTTCCATCTAAGAATTTAGGGATTTGGGTCCCTCAGATTCCACAACTCTTCTCTCT
TCATACCCAAATCAATACTTTTCTCAAGACAACAACATACAGTTTGGGATGTACAACAAA
GAAACATCACCAGAAGTAGTGTCTTTTGCTCCAATACAAAACATGAAACAACAAGGACAG
AACAACAAGAGATGGTATGATGGTGGCTTCACTGTCCACAGATCACTCCTCCTCCT
CTTTCTCTAATAAAAAAGTTTAGATCTTTCTGGTAA

>G1482 Amino Acid Sequence (domain in aa coordinates: 5-63)

MKIRCDVCDKBEASVFCTADEASLCGGCDHQVHANKLASKHLRFSLLYPSSNTSSPLC
DICQDKKALLFCQQDRAILCKDCDSSIHAANEHTKKHDFLLTGVKLSATSSVYKPTSKS
SSSSSSNQDFSVPGSSISNPPPLKKPLSAPPQSNKIQPFKINGGDASVNQWGSTSTISE
YLMDTLPGWHVDFLDSSLPTYGFSKSGDDDGVLPMPEPDNNTKRNNNNNNNNNNTV
SLPSKNLGIWVPQIPQTLPSYPNQYFSQDNNIQFGMYNKETSPFVVSFAPIQNMKQGGQ
NNKRWYDDGGFTVPQITPPPLSSNKKFRSFW*

>G225 (157..441)

CTCTCTCTCTCACTCTTTTCTTTTCCGAGAACCCAACAAAAAAGCTACTATTAATCC
TTCCCTCGTGAGGAAATCATTTCTTCTGTTTCTCGAGATTTATTCTCTTTCTCTCTCT
CTTTCTCTGTGTGTTTCGTGTCTTCAGATTAGTTTCGATGTTTCGTTTCAGACAAGCGGAA
AAAATGGATAAACGACGACGAGAGACAGAGCAAAGCCAAGGCTTCTTGTTCCGAAGAGGTG
AGTAGTATCGAATGGGAAGCTGTGAAGATGTGAGAAGAAGAAGATCTCATTTCTCGG
ATGTATAAACTCGTTGGCGACAGGTGGGAGTTGATCGCCGGAAGGATCCCGGACGGACG
CCGGAGGAGATAGAGAGATATTGGCTTATGAAACACGGCGTCGTTTTTGCCACAGACGA
AGAGACTTTTTTAGGAAATGATTTTTTTGTTTGGATTAAAAGAAAATTTCTCTCTCTT
AATTACACAAGACAAGAAAAAAGGAAATGTACCTGTCTTGAATTACTATTTTGAATGT
ATAATTATCTATATATATAAGAAGAAAAAATTGCTTAGGAATTT

>G225 Amino Acid Sequence (domain in AA coordinates: 39-76)

MFRSDKAEMDKRRRRQSKAKASCSEEVSSIEWEAVKMSEEEEDLISRMYKLVGDRWELI
AGRIPGRTPEEIERYWLMKHGVVFANRRRDFFRK*

>G226 (10..348)

CCAGTAGTTATGGATAATACCAACCGTCTTCGTCTTCGTCCGGTCCCAGTCTTAGGCAA
ACTAAGTTCACTCGATCCCGATATGACTCTGAAGAAGTGAGTAGCATCGAATGGGAGTTT
ATCAGTATGACCGAACAAGAAGAAGATCTCATCTCTCGAATGTACAGACTTGTCCGTAAT
AGGTGGGATTTAATAGCAGGAAGAGTCGTAGGAAGAAAGGCAATGAGATTGAGAGATAC
TGGATTATGAGAACTCTGACTATTTTCTCACAACGACGACGTCTTAATAATTCTCCC
TTTTTTTCTACTTCTCTCTTAATCTCCAAGAAAATCTAAAATGTAAAGAAATCAAAT
AAAAGCTTTCAATCAATAAAGTAGAACAAATCTTGAATGTCTTCTCA

>G226 Amino Acid Sequence (domain in AA coordinates: 28-78)

MDNTNRLRLRRGPSLRQTKFTRSRDYSEEVSSIEWEFISMTEQEEDLISRMYRLVGNRWD
LIAGRVVGRKANIEIRYWIMRNSDYFSHKRRRLNNSPFFSTSPNLQENLKL*

>G9 (81..1139)

GTGTTTCTTCTTTCTGCTAAAGGTTATAATTTTTGTTTCTTGGTTTGGTGAGAATCTTC
AAGAACTGAAACAAAGAAAATGGATTCTAGTTGCATAGACGAGATAAGTTCTCCACTT
CAGAATCTTTCTCCGCCACCACCGCCAAGAAGCTCTCTCCTCCTCCCGCGGCGGCTTAC
GCCTCTACCGGATGGGAAGCGGCGGGAGCAGCGTCGTGTTGGATCCCGAGAACGGCTAG
AGACGGAGTACGAAAGCTACCATCTTCAAAATACAAAGGTGTTGTTCTCAGCCTAACG

GAAGATGGGGAGCTCAGATCTACGAGAAGCACCAACGAGTATGGCTCGGGACTTTCAACG
 AGCAAGAAGAAGCTGCTCGTTCCTACGACATCGCAGCTTGTAGATTCCGTGGCCGCGACG
 CCGTCGTCAACTTCAAGAACGTTCTGGAAGACGGCGATTTAGCTTTTCTTGAAGCTCACT
 CAAAGGCCGAGATCGTCGACATGTTGAGAAAAACACTTACGCCGACGAGCTTGAACAGA
 ACAATAAACGGCAGTTGTTTCTCTCCGTCGACGCTAACGGAAAACGTAACGGATCGAGTA
 CTACTCAAAACGACAAAAGTTTTAAAGACGTGTGAAGTCTTTTCGAGAAGGCTGTTACAC
 CTAGCGACGTTGGGAAGCTAAACCGTCTCGTGATACCTAAACAACACGCCGAGAAACACT
 TTCCGTTACCGTACCGTACCGGCAGTGAATAAGGAGTTTGTATCAACTTCGAAGACG
 TTAACGGTAAAGTGTGGAGGTTCCGTTACTCATACTGGAACAGTAGTCAAAGTTACGTGT
 TGACCAAGGGATGGAGTCGATTTCGTCAAGGAGAAGAATCTTCGAGCCGGTGATGTTGTTA
 CTTTCGAGAGATCGACCGGACTAGAGCGGCAGTTATATATTGATTGGAAGTTCGGTCTG
 GTCCGAGAGAAAACCCGGTTCAGGTGGTGGTTCGGCTTTTCGAGTGTGATATCTTTAATG
 TGACCACCGTGAAGCCAAACGACGTCGTGGCCGTTTTCGGTGGAAAGAGATCTCGAGATG
 TTGATGATATGTTTTCGCTTACGGTGTTCGAAGAAGCAGGCGATAATCAATGCTTTGTGAC
 ATATTTCTTTTCCGATTTTATGCTTTCGTTTTTAATTTTTTTTTTTTGTCAAGTTGTGT
 AGGTTGTGATTATGCTAGGTTGTATTTAGGAAAAGAGATAAGACC

>G9 Amino Acid Sequence (domain in AA coordinates: 62-127)
 MDSSCIDEISSSTSESFSAATAKKLSPPPAALRLYRMGSGGSSVLDPENGLTESRKL
 PSSKYKGVVPQPNRWGAQIYEKHQRVWLGTFFNEQEEAARSYDIAACRFRGRDAVNVFN
 VLEDGDLAFLEAHSKAEIVDLRKHTYADELEQNNKRLFLSVDANGKRNGSSTQNDKV
 LKTCFVLFKAVTPSDVGKLNRLVLPKQHAKEKHFPLPSPSPAVTKGVLFEDVNGKVWR
 FRYSYWNSSQSYVLTGWSRFVKEKNLRAGDVVTFERSTGLERQLYIDWKVRS GPENPV
 QVVVRLFGVDIFNVTTVKPNDVVAVCGGKRSRDVDDMFALRCSKKQAIINAL*

>G1040 (51..863)
 CTTTGATCTCCACTATTTAAGTAGACAAGAATCATAAAGAAAATAGTGAGATGATGATGT
 TAGAGTCAAGAAACAGTATGAGAGCTTCAAACTCAGTCCCAGATCTGTCTCTTCAGATCA
 GTCTTCTTAACATATCACGCCGAAAACCTCTTCACGGCGGTGACCGGAGCTCCACAAGCA
 GTGATTCTGGAAGCAGCCTCAGTGACCTGAGCCATGAGAACAACCTCTTCAACAAACCTC
 TCTTGAGCTTAGGATTTGACCATCATCATCAAAGGCGCTCAAACATGTTCCAACCTCAAA
 TCTACGGTCGAGATTTCAAGAGAAGCTCATCATCAATGGTTGGTCTTAAACGAAGCATTCT
 GTGCTCCAAGAATGAGATGGACTTCTACTCTTCATGCTCACTTCGTCATGCTGTTCAAC
 TTCTTGGCGGCCATGAAAGAGCAACGCCATAATCAGTGTGGAGCTCATGAATGTGAAGG
 ATCTAACCTTAGCTCATGTCAAGAGTCACTTGACAGATGTATAGAACAGTGAAATGCACTG
 ATAAAGGATCACCAGGAGAAGGAAAGGTAGAGAAAGGCAGAGCAGAGGATAGAGGACA
 ATAATAATAATGAAGAGCTGATGAAGGAAGTACACAAATTCGCCAAACTCATCATCTG
 TGCAAAAGACCCAAAGAGCTTCATGGTCATCGACAAAGGAAGTATCTAGGAGCATATCTA
 CACAAGCATATTCTCACTTGGGAACAACATCATCACTAAGGCCAATGAAGAGAAAGAGG
 ATACCAACATTTCATCTCAATTTGGATTTACATTGGGCGGCCCTAGTTGGGGGATGGAATA
 TGCGGAACCTCCAGTGATTTAACCTTCTCAAGTGCTAATTGCCTTAAGCTACAACAAA
 TAAGTCAGCTTAGGTTACCAGTTTAAACATAATTTTAACTTGTTTTGATCATATGAGCTT
 CGGAAGAATCATATTATCATCATATATGAACCTCTTCCAAGAATGTCTATGAGTTTTT
 TGATATGTATAATCAAGAGAATCGTTTGAAGTAAAAA

>G1040 Amino Acid Sequence (domain in AA coordinates: 109-158)
 MMMLESRNSMRASNSVPLSLQISLPNYHAGKPLHGGDRSSTSSDSGSSLSDLSHENNFF
 NKPLLSLGLFDHHQRRSNMFQPIYGRDFKRSSSMVGLKRSIRAPRMRWTSTLHAHFVH
 AVQLLGGERATPKSVLELMNVKDLTLAHVKSHLQMYRTVKCTDKGSPGEGKVEKEAEQR
 IEDNNNNNEADEGTDINSFNSSSVQKTQRASWSSTKEVSRISISTQAYSHLGTTHHTKANE
 EKEDTNIHLNLDFTFGLVGGWNMRNPPI*

>G2114 (64..1311)
 ATAAAACGAAACCTATACATATAAACTAAGAGCGAGAAAGACAGCTAGAGAGAGAGAGA
 GAGATGAAGAAATGGTTGGGATTTTCAATTGACACCTCCTTTGAGAATCTGCAATAGTGAA
 GAAGAAGAACTTAGGCATGACGGTTCGATGTTTGGAGATATGATATTAACCTTTGATCAT
 CATCATCATGATGAAGACGTTCCAAAGGTGGAAGATCTCCTCTCAAACCTCTCATCAAACC
 GAGTATCTATAAACCATAAACCAACCAATGTCAACTGCACCACTGTGGTTAACAGGTTA
 AACCACCCCGTTACCTTCTCCACGACCAACCGTAGTTACACCACATTACCCGAACCTA
 GATCCGAACCTTAGCAATGATTATGGAGGTTTGGAGGGTTCGGTTCGGTCTCGGTTTTTC
 AAATCTTGGTTAGAGCAAGGCACTCCAGCATTCCCACTCTCGAGTCAATACGTTACTGAA

GAGGCTGGTACGAGCAATAATATTAGTCATTTTAGTAACGAAGAGACTGGTTATAACACC
AATGGCTCAATGCTATCATTGGCTTTGAGCCATGGGGCTTGTCTGATTTGATCAACGAA
TCGAATGTATCCGCACGGGTCGAAGAACCGGTTAAGGTAGATGAGAAGCGGAAGAGATTG
GTTGTTAAACCTCAGGTAAAGGAATCCGTTCTCGGAAGTCGGTTGATAGTTATGGACAA
AGAAGTTCTCAGTATCGTGGAGTTACAAGGCATAGATGGACAGGGAGATATGAAGCTCAC
TTATGGGATAATAGCTGTAAGAAGGAGGGACAGACAAGGAGAGGAAGACAAGTGTATCTT
GGAGGGTATGATGAGGAGGAGAAAGCAGCGAGGGCATATGATTTAGCGGCTCTGAAGTAT
TGGGGTCTTACCCTCACTTAAATTTCCCTTTGAGTAATTACGAAAAGGAGATCGAGGAA
CTCAATAACATGAATCGGCAAGAATTTGTTGCCATGTTGAGGAGGAATAGCAGCGGGTTT
TCGAGGGGAGCTTCCGTGTATAGAGGAGTTACAAGGCATCATCAACATGGAAGGTGGCAA
GCCAGAATTGGAAGAGTTGCTGGAAACAAGGACTTGTACCTTGAACATTTAGCAGCGAA
GAAGAAGCAGCGGAGGCGTACGATATCGCGGCAATTAAATTCAGAGGCCTAAACCTGTGA
ACCAATTTTCGATATAAATAGATATGACGTGAAGAGGATATGTTCAAGCTCAACGATTGTT
GATAGCGACCGAGGCCAAACATTCTCCACAGCTCTGGCGCCGCCACTAACCACACCG
TAAACTCCTCGCCGGAGAGACTATTCCACAGCTACGGTTGGTTGAGGAAATAAGTTCGTC
CAGTCTGTTTAAATCATTTATGGTTTAAATAACATATATTCCTAAGTAATTGAGGCCGGTC
TACATATATACAACTTTTTTAGCAAATTAAGTTATCAGAATCCACTATATATTATTCTCT
>G2114 Amino Acid Sequence (conserved domain in AA coordinates:221-297, 323-393)

MKKWLGLFSLTPPLRICNSEEEELRHDGSDVWRYDINPDHHHDEEDVPKVEDLLNSHQTE
YPINHNQTNVNCCTTVNRLNPPGYLLHDQTVVTPHYPNLDPNLSNDYGGFERVGSVSVFK
SWLEQGTAPFPLSSHVTEEAGTSNNISHFSNEETGYNTNGSMLSLALSHGACSDLINES
NVSARVEEPVKVDEKRLVVKPQVKESVPRKSVDSYGQRTSQYRGVTRHRWTGRYEHL
WDNSCKKEGQTRRGROVYLGYYDEEKAARAYDLAALKYWGPTTHLNFPLSNYEKEIEEL
NNMNRQEFVAMLRNSSGFSRGASVYRGVTRHHQHGRWQARIGRVAGNKDLYLGTFTSQE
EAAEAYDIAAIKFRGLNAVTFNFDINRYDVKRICSSSTIVDSQAKHSPTSSGAGH*
>G450 (65..751)

GAGTTATCGAGAGAGAGAGAAAAATATTTCTGATTTAAGACATATATAGACAGCAAGAAG
AGATATGAACCTTAAGGAGACGGAGCTTTGTCTTGGCCTCCCCGGAGGCACTGAAACCGT
TGAAAGTCCGGCCAAGTCGGGTGTTGGGAACAAGAGAGGCTTCTCCGAGACCGTTGATCT
CAAACCTTAATCTTCAATCTAACAAACAAGGACATGTGGATCTCAACACTAATGGAGCTCC
CAAGGAGAAGACCTTCTTAAAGACCTTCTAAGCCTCCTGCTAAAGCACAAAGTGGTGGG
TTGGCCACCGGTGAGGAACCTACCGGAAAAATGTTATGGCTAATCAGAAGAGCGGCGAAGC
AGAGGAGGCAATGAGTAGTGGTGGAGGAACCGTCGCCTTTGTGAAGGTTTCCATGGATGG
AGCTCCTTATCTTCGGAAGGTTGACCTCAAGATGTACACCAGCTACAAGGATCTCTCTGA
TGCTTTGGCCAAAATGTTTCAGTCTCTTTACCATGGGGAGTTATGGAGCACAAGGGATGAT
AGATTTTCATGAACGAGAGTAAAGTGATGGATCTGTTGAACAGTTCTGAGTATGTTCCAAG
CTACGAGGACAAAGATGGTGAAGTGGATGCTCGTTGGTGATGTCCCTTGCCGATGTTTGT
CGAGTCATGCAAAACGTTTGCGCATAATGAAAGGATCCGAAGCAATTGGACTTGCTCCAAG
AGCAATGGAGAAGTTCAAGAACAGATCATGAACAAAAAAGAGGACAAATATGCATTG
ATTTTTTTTTTTTTTTGGTATTGTTATGATCATGTGTTTAAATTTAAATATAGGAAGGATA
TAGGAAAAATATAATTGTTTACAAAAAATAACTTTAAATATGTCTTTTTTTTTTTTTTTGA
AATTAGTCTGTGTTTTTGTTCATCTCTTAATTAGTAGAAATCATTTTTTAAATATGTAA
TTGTGATAGTAAATCTATAGAGTTTCGTA

>G450 Amino Acid Sequence (domain in AA coordinates: TBD)
MNLKETELCLGLPGGTETVESPAKSGVGNKRGFSETVDLKLNLQSNKQGHVDLNTNGAPK
EKTFLKDPSPKPPAKAQVVGWPPVRNRYRKNVMANQKSGEAEEMSSGGGTVAFVKVSMGDA
PYLRKVDLKMYSYKDLSDALAKMFSSFTMGSYGAQGMIDFMNESKVMDDLNSSEYVPSY
EDKGDWMLVGDVPWPMFVESCKRLRIMKGSSEAIGLAPRAMEKFKNRS*

>G584 (40..1809)
AAAAAGTCTTCTCTTTTATAACTACGTCAGAGAACTGTTATGTCTCCGACGAATGTTCAA
GTAACCGATTACCATCTCAACCAATCAAAAACGGATACAACAAATCTCTGGTCAACCGAC
GACGATGCATCGGTAATGGAAGCTTTCATCGGCGCGGCTCCGATCATTCTTCTTTTTT
CCTCCACTTCTCTCTCTCTCTCTCTCAAGTCAACGAAGATAATCTCCAGCAACGTCTC
CAAGCTTTAATCGAAGGAGCAAACGAGAACTGGACTTACGCCGTGTTCTGGCAATCATCT
CACGGTTTCGCCGGAGAAGACAACAACAACAACACAGTGTGTTAGGTTGGGGAGAT
GGTTATTACAAAGGAGAAGAAGAGAAGTCTAGAAAGAAGAAATCAAATCCAGCTAGTGCA
GCTGAACAAGAGCATCGTAAGAGAGTGATTAGAGAGCTCAACTCTTAAATCTCCGGTGGT

GTAGGAGGAGGAGATGAAGCTGGAGATGAAGAAGTTACAGATACTGAATGGTTCTTCTTA
GTTTCAATGACACAGAGCTTTGTCAAGGGTACTGGTTTACCTGGTCAAGCTTTCTCAAAT
TCAGACACGATTTGGTTATCTGGTTCTAATGCTTTAGCTGGATCAAGTTGTGAGAGAGCT
CGTCAAGGTCAGATTTATGGGTTACAAACAATGGTGTGTGTAGCGACAGAGAATGGTGTG
GTTGAGCTTGGTTCGTCGGAGATTATTCATCAAAGTTCAGATCTTGTGTATAAAGTTGAC
ACCTTTTTCAATTTTAACAATGGTGGTGGTGAATTTGGTTCTTGGGCGTTTAATTTGAAT
CCAGATCAAGGAGAGAATGATCCAGGTTTGTGGATTAGTGAACCTAATGGTGTGACTCT
GGTCTTGTAGCTGCTCCGGTGATGAATAATGGTGGAAATGACTCAACTTCTAATTTCTGAT
TCTCAACCAATTTCTAAGCTTTGTAAATGGAAGCTCTGTTGAAAACCTTAACCTAAAGTT
CTGAAATCTTGTGAAATGGTGAATTTCAAGAATGGGATTGAGAATGGTCAAGAAGAAGAT
AGTAGTAATAAGAAGAGATCACCGGTTTCAATAATGAAGAAGGGATGCTTTCTTTTACC
TCTGTTCTTCCATGTGACTCGAATCACTCTGATCTTGAAGCTTCAGTGGCTAAAGAAGCT
GAGAGTAACAGAGTTGTGGTTGAACCGGAGAGAAGAAACCGAGGAAACGAGGGAGAAAACCG
GCGAATGGAAGAGAAGAGCCTTTGAATCATGTAGAGGCAGAGAGACAGAGAAGAGAGAAG
TTGAATCAGAGATTCTATTCTTTAAGAGCTGTGGTTTCTAATGTGTCTAAGATGGATAAA
GCTTCTCTATTAGGAGATGCTATTTCTGATATCAGTGAGCTTAAGTCTAAGTTGCAAAAG
GCTGAATCTGATAAAGAAGAGTTGCAGAAGCAGATTGATGTGATGAATAAAGAAGCGGGA
AATGCCAAAAGTTCCGGTAAAGATCGAAAATGTTTGAATCAAGAATCGAGTGTGTTGATA
GAGATGGAGGTTGATGTGAAGATTATTGGTTGGGATGCAATGATAAGGATTCAATGTAGT
AAGAGGAATCATCCTGGTGTCTAAGTTCATGGAAGCACTTAAGGAGTTGGATTGGAAGTG
AATCATGCGAGTTTATCGGTAGTGAATGATCTTATGATCCAACAAGCGACTGTGAAAATG
GGGAATCAGTTTTTTCACGCAAGATCAACTCAAGGTTGCTCTAACGGAGAAAGTTGGAGAA
TGTCCATGAATTGAAGTCAGCATCTTTAGGGCTAATACACCGGAGAATACTGCGAAAAGT
CGAAAACAACGATCATAGTATAAGCCGCGGTAAAAAGTGTAAACCTTTACACAAGTTT
CTCTAGTGAATGTAGTTGTAACTCTATTGTGTAAAGGTAATTTTGTAGTACCCACTTGT
TGCTATTGAATGCTTGTGTAGAGGATTCTTAGTGTAGTATATGATTAGGTTGGGGTTTG
TTGTTTCATGAGATAAATAAATGTGTTTGATCAATGGTTAAGTCTTTGGTTTGTGGTGT
ATGTATGTAAATAAGGCTTTTGTAGAAATAAGACAAATGGGACTGAAGTTGGAGTTTAA
AA

>G584 Amino Acid Sequence (domain in AA coordinates: 401-494)
MSPTNVQVTDYHLNQSKTDTTNLWSTDDDDASVMEAFIGGSDHSSLFPPLPPPLPQVNE
DNLQQLQALIEGANENWTVYVFWQSSHGFGEDNNNNNTVLLGWGDGYKGEEEKSRKK
KSNPASAAEQEHRKRVIRELNSLISGGVGGGDEAGDEEVDTEWFFLVSMQSFVKGTGL
PGQAFSNSDTIWLSGSNALAGSSCERARQGQIYGLQTMVCVATENGVVVELGSSEIIHQSS
DLVDKVDTFNFFNNGGGEFGSWAFNLNPDQGENDPGLWISSEPNVDSGLVAAPVMNNGN
DSTSNSDSQPI SKLCNGSSVENPNPKVLKSCMVNFKNIGENGQEESSNKKRSPVSNNE
EGMLSFTSVLPDCSNHSDLEASVAKEAESNRVVVEPEKKPRKRGRKPANGREEPLNHVEA
ERQRREKLNQRFYSLRAVVPNVSKMDKASLLGDAISYISELKS KLQKAESDKBELQKQID
VMNKEAGNAKSSVKDRKCLNQESSVLIEMEVDVKIIGWDAMIRIQSKRNHPGAKFMEAL
KELDLEVNHASLSVNDLMIQQATVKMGNQFFTQDQLKVALTEKVGEC*

>G668 (1..1056)

ATGGGAAGACCCTTGCTGTGAAAAGATTGGAGTGAAGAAAGGGCCATGGACACCAGAG
GAAGACATCATCTTGGTTTCTTACATCCAAGAACATGGTCTGGAACCTGGAGATCTGTC
CCAACACACACAGGTTTAAGATGTAGCAAGAGCTGCAGATTGAGATGGACTAATTATCTT
CGACCCGGTATTAAGCGTGGAATTTTACTGAGCATGAAGAGAAGACAATTGTTTCATCTT
CAAGCCCTTTTAGGCAACAGATGGGCAGCCATAGCATCATACCTTCCAGAAAGGACAGAC
AATGATATAAAGAACTATTGGAACACTCACTTGAAGAAGAAGCTCAAAAAGATTAATGAA
TCTGGTGAAGAAGATAATGATGGTGTCTCTTCATCAAACACTAGTTTCACAAAAGAACCAT
CAAAGCACTAACAAGGTCAATGGGAAAGAAGACTTCAGACAGACATTAACATGGCAAAA
CAAGCTCTTGTGAGGCCTTGTCTTTAGACAAAACCATCATCCACTCTTTCATCATCTTCA
TCATTACCGACACCAAGTAATCACACAACAAACATCCGTAACCTTCTCATCAGCTTTGCTT
GACCGTTGTTATGATCCATCCTCTTCTTCTTCTTACCTACCAACCACTACCAAGCAAC
ACTACTAATCCATACCCATCAGGGGTATATGCGTCAAGTGCTGAGAACATCGCCCGGTTG
CTTCAAGATTTTCATGAAAGACACACCCAAGGCTTTAACTTTATCATCTTTCATCTCCGTT
TCAGAGACTGGACCACTCACTGCTGAGTCTCGGAAGAAGGTGGAGAAGGGTTTGAACAA
TCTTTCTTCAGCTTCAATTCATGGACGAAACTCAAACTTGACTCAGGAGACAAGCTTC
TTCCATGATCAAGTGATCAAACCGGAAATAACAATGGACCAAGATCATGGTCTAATATCA

CAAGGGTCTCTGTCTTTGTTTGAGAAATGGTTATTTGATGAGCAAAGCCACGAGATGGTT
GGTATGGCACTAGCAGGACAAGAAGGGATGTTCTAG
>G668 Amino Acid Sequence (domain in AA coordinates: 13-113)
MGRPPCCEKIGVKKGPWTPEEDIILVSYIQEHGPGNWRVPTHTGLRCSKSCRLRWNYL
RPGIKRGNFTEHEEKTIVHLQALLGNRWAAIASYLPERTDNDIKNYWNTHLKKLKKINE
SGEEDNDGVSSSNTSSQKNHQSTNKGQWERRLQTDINMAKQALCEALSLDKPSSLSSSS
SLPTPVITQQNIRNFSSALLDRCYDPSSSSSTTTTTNTNPNYPSPGVYASSAENIARL
LQDFMKDTPKALTLSSSSPVSETGPLTAAVSEEGGEGFEQSFFSFNSMDETQNLQTQTSF
FHDQVIKPEITMDQDHGLISQGSLSLFEKWLFEQSHMVMALAGQEGMF*
>G1050 (23..1582)
TTCCCCATTTTCAGAAAATCAAAATGGGTGGTGGTGGTGATACAACAGATACCAATATGAT
GCAGAGAGTTAATTCTTCTTCTGGTACATCGTCTTCTTCGATCCCTAAACACAATCTTCA
CTTGAATCCTGCTCTTATCCGCTCTCACCATCACTTCCGTCAACCTTTTACCGGAGCTCC
TCCACCGCCGATTCCACCCATTTCTCCTTACTCTCAGATCCCGGCGACTTTACAACCTAG
ACATTCTCGCTCTATGTCGAACCGTCTTCTTCTTCTCCTTTGATTTCATTGCCGCCGT
AAATCCTTCTGCTCCGTCGGTTTCGGTGTTCGGTGGAGGAGAAAACCGGTGCCGGATTAG
TCCCTTCGTTGCCCTCCGTCAACCGTTTACGATGTGTCTTCTTAGCTCTAGGAACGCCGG
AGATGGAGAGAATCTACCTCCGAGAAAGTCGCATAGGCGTTCGAATAGTGATGTTACTTT
TGGGTTTAGTTCAATGATGTCTCAGAATCAAAAGTCTCCTCCTTTGAGTTCTTTGGAGAG
ATCGATCTCTGGTGAAGATACATCAGATTGGTCTAATTTGGTGAAGAAAGAACCGAGAGA
AGGCTTCTACAAGGGAAGAAAACAGAGGTTGAAGCAGCTATGGACGATGTTTTCACGGC
TTATATGAATCTTGATAACATTGATGTCTTGAATTCTTTTGGAGGTGAAGATGGCAAGAA
TGGGAATGAGAATGTGGAGGAGATGGAGAGTAGTAGAGGTAGTGGTACAAAGAAGACGAA
TGGTGAAGTAGTAGTGATTCTGAAGGAGATAGCAGTGCAGTGGGAATGTGAAGGTTGC
GTTGAGTTCTTCTTTCAGGCGTGAAGAGAAGAGCAGGTGGAGATATTGCTCCTACTGG
TAGACATTACAGGAGTGTCTTCTATGGACAGTTGTTTCATGGGGAAGTTGAATTTCCGCCA
CGAATCATCGCTAAAGCTTCCGCCCTTCTTCATCAGCTAAAGTTTCCCCAACCAATTCAGG
TGAAGGGAATTCAAGTGCTTATAGTGTGAATTTGGAACAGTGAGTTTACTGCAGCTGA
AATGAAGAAGATTGCAGCTGATGAGAACTCGCTGAGATTGTAATGGCTGACCCTAAGCG
TGTTAAAAGAATCTTGGCGAACC CGGTATCTGCTGCACGTTCAAAGGAGCGGAAGACGCG
ATACATGGCAGAGTTGGAACACAAGGTGCAGACACTTCAGACTGAAGCTACTACATTATC
GGCTCAGCTCACACATTTGCAGAGAGATTCTATGGGGTTGACAAACCAGAACAGTGAGCT
GAAGTTTCGTCTTCAAGCTATGGAGCAGCAAGCACAACCTCCGCGATGCTCTGTGAGAGAA
ACTGAATGAAGAAGTCCAGCGTTGAAACTGGTGTAGGGGAGCCGAACCGCAGGCAAAG
TGGGAGCAGCAGCAGCGAATCAAAGATGTCACTAAACCCGAGATGTTTTCAGCAGCTTAG
CATAAGTCAGTTACAACACCAACAGATGCAGCATTCCAATCAGTGTAGCACAATGAAAGC
AAAGCACACTTCAAACGACTAGGGTAAGTAAACTGCGATCCGCAGTTGTCTAGTTACAT
ATATGATAAGAATCTTTTGTGAGAGTTCTGTTTTTGAAGTTTTTAAAGAAACATATATA
AAGATTATGTCCGGGAAATTTGATCATATTTCTGAAACATACACATATATATAGTGG
TAATGGAGGACTTTCTTTCTGGACCA
>G1050 Amino Acid Sequence (domain in AA coordinates: 372-425)
MGGGGDTTDTNMMQVRVNSSSGTSSSSIPKHNHLNLPALIRSHHHFRHPFTGAPPPPPIPI
SPYSQIPATLQPRHSRSMSPSSFFSFDLPLNPSAPSVSVSVEEKTGAGFSPSLPPSP
FTMCHSSSSRNAGDGENLPPRKSHRRSNSDVTFGFSSMMSQNKSPPLSSLERSISGEDT
SDWSNLVKKEPREGFYKGRKPEVEAAMDVFTAYMNLNIDVLNSFGGEDGKNGNENVEE
MESSRSGTKKTNNGSSSDSEGDSSASGNVKVALSSSSSGVKRRAGGDIAPTGRHYSVS
MDSCFMGLNFGDESSLKLPPSSSAKVSPINSGEENSSAYSVEFGNSEFTAAEMKKIAD
EKLAEIVMADPKRVKRI LANRVSAARSKERKTRYMAELEHKVQTLQTEATTLQAQLTHLQ
RDSMGLTNQNSLKFRLQAMEQQAQLRDALSEKLNVEVQRLKLVIGEPNRRQSGSSSSSES
KMSLNPMEFQQLSISQLQHQMHSNQCSMTMAKHTSND*
>G1463 (199..1209)
TATCCTTCGCAAGACCCCTTCTCTATATAAGGAAGTTTCATTTTCATTTGGAGAGGACACGC
TGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAGTTTGAGATTTGCTTCATCCGGT
TTTTTTATTTTCTGCAAAATATGTCACTCTCTCCATTTTGTTCATATATAATATGTTG
AAGTTTGATCAACTTAGTATGCGTTCTTTTCTCTCTAGTTCTCTGTTTCTTGGTCTGA
TTTAGTTTCGTTATGGCGGACACACTGCTCAACGCAGAGACGAAGTAATAATCTCACGT
TATCTGAAGCCTATGATCGTTAACAGAGTATCATGGCCTGATCTCTTCATCGAAGACGCA

GACGTGTTCAACAAGGATCCATATGTGAAGTTCATGCTGAGATCCCTAGCTTCGTGATC
 GTTAAACCACGAACAAAGGCTTGTGGTAAAACCGATGGATGTGATTCCGGTGTGCGGAGG
 ATCATTGGTCGTGATAAGCTGATAAAGTCGGAGGAGACTGGTAAGATTCTAGGGTCAAG
 AAGATACTCAAGTTCTGCCTAAAGTGGAAACCTAGAGAATACAAGAGAAGTTTGGTAATG
 GAAGAGTATAGGCTTACCAATAACTTCAACTGGAAGCAAGATCATGTGATTGCAAGATT
 CGGCTTTTGTGTTGAAGCAGAAATTAGTTTCTTGCTAGCCAAGCATTCTTACACTACATCA
 GACTCACTTCCTCGAAATGTGCTGTTGCCAGCTTATGGATTCTGTTTACCAGATAAACAA
 GAGGAGGACGAATTTATCCGGTGACGATAATGATTTTCAAGGAAAAGATTGGCCTAGC
 TACGTTACCAACAACGTGTATTGTTCTGCATCCATCGGAGCTTGTGAATGTTACGATGGG
 AAGTTTCATGATAACGGAATCTGCATCTTCGCTAACAGGACTTGTGGTGTAAACCGATAAA
 TGCAATGAAGGTTACTGGAAGATTAAGCACCGTGAGAAGCTGATCATGTACGGTACGGG
 CAGACCATTGGTTGGAAGAAAGTTTTTTCAGTTTATGAAACGGAGAAAGAAAGACATTTT
 GGTAATGGAGAAGAAGTGAAGGTAACCTGGACTCTAAAAGAGTATAGGCTTACCAGAAAA
 ATGAACAAGAATAAAGTGGTGTGCGTTATCAAGTATAAGGTAAAGTGTTTACCGAGGATA
 ACTAGCTAGGGACTTCTACTCTTGGTTTCATGATCGATCGACCGCTCTAGACAGGCCTC
 GTACCGGATCCTCTAGCTAGAGCTTTCTGTTCTGATCATCGGTTTCGACAACGTTTCGTCA
 >G1463 Amino Acid Sequence (conserved domain in AA coordinates:9-156)
 MRFFSLVPLFLGRFSFVMADTLLNAEDEVIISRYLKPMIVNRVSWPDFIEDADVFNKD
 PYVKFHAELPSFVIVKPRTKACGKTGDCDSGCWRIIGRDKLIKSEETGKILGFKKILKFC
 LKWKPREYKRSVMEEYRLTNFNWQDHVICKIRLLFEAEISFLLAKHFYTTSDSLPRN
 VLLPAYGFCSPDKQEEDEFYPTIMISEGKDWPSYVTVNNVYCLHPSSELVNVHDGKFHDNG
 ICIFANRTCGVTDKCNENYWKIKHREKLIMSRYGQTIGWKKVFQFYETEKERHFGNGEEV
 KVTWTLKEYRLTRKMNKNKVVCVIKYKVKCLPRITS*

>G1944 (236..1306)
 TCGACCTTCCTAATTTCCAACCTCTGTTCTTAGCAATATATTTTTTCTCCAAAAATAATT
 CTCAGTTTGATTTTCTTCTTCTAGCTCTTAAGTATATTTCTTTGTTGTTATTTATCTTTT
 AATCCTTTAATCTCATCTTTGTTTATCTTTAATCAAAACCCAAAATTTACATGGGTCTT
 GAAATCTAGAAGAAATAAAGGAAACATAACAAAAATAGAAAGAAAAAGAGCTAATGGT
 CTTAAATATGGAGTCTACCGGAGAAGCTGTTAGATCAACCACCGGTAACGACGGTGGTAT
 TACGGTGGTTAGATCCGACGCGCGCTCAGATTTCCACGTAGCTCAAAGATCAGAAAGCTC
 AAACCAATCTCCACCTCTGTCACTCCTCCTCCACCACAGCCATCGTCTCATCACACAGC
 TCCTCCGCGCTGCAAATTTGACGGGTGACGACTACGACTACGACGGCCCGGATGGAAGG
 TATCTCCGGTGGACTGATGAAGAAGAAGCGTGGACGGCCAAGGAAGTATGGACCGGACGG
 GACTGTTGTAGCGTTATCTCTTAAACCGATTTCATCAGCGCCGGCGCGCTCGCATCTTCC
 GCCGCGAGTTTACACGTCATCGATTTCTCCGCTTCTGAGAAACGTAGCAAAGTGAAACC
 AACGAACTCGTTTAAACAGAACAAAGTATCATCACCAAGTTGAGAATTTGGGTGAATGGGC
 TCCTTGCTCCGTCGGTGGTAATTTACACCTCATATAATCACAGTCAACACCGGCGAGGA
 TGTAACAATGAAGATAATCTCGTTTTCGCAACAAGGACCTCGCTCTATTTGTGTTCTGTC
 AGCAAACGGTGTATTTCAAGCGTTACACTTCGTGACCCAGATTCCTCTGGCGGCACATT
 GACATACGAAGGTCGGTTTGAGATATTATCATTATCCGGGTCATTATGCCTAATGATTC
 AGGCGGAACACGAAGTAGAACGGGAGGAATGAGTGTATCGTTAGCAAGTCCCGATGGACG
 TGATAGTAGGCGGTGGCCTCGCCGGTTTACTAGTAGCCGCGAGTCCGGTTTCAAGTGGTGT
 AGGAAGTTTTTTAGCGGGCACTGACCATCAAGATCAGAAACCGAAAAAGAACAAACATGA
 TTTCATGTTGTGTCGAGTCTTACCGCTGCAATTCCTATCTCTAGTGCAGCTGATCACCGGAC
 AATCCATTCCGTCTCGTCTCTTCCGGTCAATAATAATACATGGCAGACTTCTTTAGCTTC
 CGATCCAAGAAACAAGCATACCGATATTAATGTCAATGTAACCTGAAATCCAATCTTTCT
 CTGTATTTTCTGTTAAACAAGTTTGATTTGGTTGTTTATCTACATAGGATTTTACTAAAA
 TGGTAGTATTATTTATAGGGTTTTAGGGTCTTTATTTTGGTTCCACTGTTGTCACTTGT
 GGATA

>G1944 Amino Acid Sequence (domain in AA coordinates:87-100)
 MVLNMESTGEAVRSTTGNDGGITVVRSDAPSDFHVAQRSESSNQSPSVTPPPPPQSSHH
 TAPPPLQISTVTTTTTTAAAMEGISGGLMKKKRGRPRKYGPDGTVVALSPKPISSAPAPSH
 LPPPSHHVIDFSASEKRSKVKTNSFNRTKYHHQVENLGEWAPCSVGGNFTPHIITVNTG
 EDVTMKIISFSQQGPRISICVLSANGVISSVTLRQPDSSGGTLTYEGRFEILSLSGSFMPI
 DSGGTRSRRTGGMSVSLASPDGRVVGGLAGLLVAASPVQVVVGSFLAGTDHQDQKPKKNK
 HDFMLSSPTAAIPISSAADHRTIHSVSSLPVNNNTWTQSLASDPRNKHTDINNVNT*
 >G2383 (37..990)

GACCTCTTTGATCCCTTCATTCCCCATCAAACAACCATGTTTCCTTCTTTTCACTACTCAC
ATTCAAAGCCCTAATTCTCACCATCACTACTCTTCGCCTTCTTTTCCCTTCTCTCCGAT
TTTCTTGAGAGTTTTGATGAATCCTTCTTGATAAACCAATTCTTGTTACAGCAGCAAGAT
GTAGCAGCAAATGTTGTTGAATCTCCTTGGAATTTTGCAAGAAGCTTGAGCTTAAGAAG
AAGAATGAGAAGTGTGTTGATGGAAGCACCTCACAGAGGTTCAATGGAGAAGGACGGTC
AAAAAAGGGACAGGCATAGTAAGATCTGCACGGCTCAAGGTCCTAGAGACCGAGGATG
AGGCTGTCTCTTCAGATTGCTCGCAAGTTTTTCGATCTTCAAGACATGTTGGGTTTCGAC
AAGGCGAGCAAGACGATTGAATGGCTTTTCTCCAATCAAAGACTTCCATCAAACAACCTT
AAAGAAAGAGTGGCTGCATCGGAAGGAGGAGGAAAGGATGAACATCTCCAGGTTGATGAA
AAGGAAAAGGATGAGACACTGAAGTTGAGAGTCTCAAAGAGAAGAACAAGACTATGGAG
AGCTCTTTTAAGACTAAAGAGTCGAGAGAGAGAGCTAGAAAGCGAGCAAGAGAGAGAACA
ATGGCAAAGATGAAGATGAGATTATTTGAGACCTCGGAAACAATTTAGATCCTCATCAA
GAACTAGAGAGATCAAGATAACCAATGGTGTACAATTACTAGAAAAGGAAAATAAAGAA
CAAGAATGGAGTAATACTAATGATGTTTACATGGTAGAGTATCAAATGGATTCTGTGAGC
ATCATAGAGAAGTTTCTTGGACTAACCAGTGACTCTAGCTCCTCTTCCATTTTTGGTGAC
TCCGAGGAATGTTACACAAGTCTTAGTTTCAAGAGGTACAATTTTCCAGCAGCAGGTAAC
AGCAATGTGTTAACTAAAACCCTAATTGAGTAATGCAGTTTTGATTAATATTAGCTTTT
TGGTAATTCCAGGAATGTCGACACCAAGGG

>G2383 Amino Acid Sequence (conserved domain in AA coordinates:89-149)

MFPSFITHIQSPNSHHYSSPSFPFSSDFLESFDESFLINQFLLQQQDVAAENVVESPWKF
CKKLELKKKNEKCVDGSTSQEVQWRRTVKKRDRHSKICTAQGPRDRMRRLSLQIARKFED
LQDMLGFPDKASKTIEWLFSKSKTSIKQLKERVAASEGGGKDEHLQVDEKEKDETLKLRVS
KRRTKTMESSEFKTKESRERARKRERERTMAKMKMRLFETSETISDPHQETREIKITNGVQ
LLEKENKEQWSENNDVHMVEYQMDSVSIIKFLGLTSDSSSSSIFGDSEECYTSLSVR
GTISAAGNSNVLTKNPN*

>G571 (326..1708)

TAGCCGACCTCTCTTCTCTTCTTGAAAAAACACCAAAGGAGCTTTAAATGCTCCGTTA
CATAATCTCTATCTCTTTCCAAGAATATAGAGAAAGGAAAATAATATACAAGAATTAAAA
GAAGGTATATCATCATCTCTCTAGCTAGTGATCAAAGCACCGTCATCATCATCATATATC
ATCAGCTTGCCCTCAGAGGAGAAGACCAACATAAGAGAGATCGAAGATCAAAATCTATCTC
TCTTCATCATCTTCTGCTGTTACTATCATATCACACGCTCTCTCAAACATCATCTTATAT
ATAGACTTCTCTTCATCATCATCAAATGCAAGGTCATCACCAGAATCATCATCAACACTT
ATCATCATCTCCGCCACGCTCTTCCATGGAACTTCATGAACAAAGATGGGTATGATAT
TGGAGAGATAGACCCCATCTTCTCTCTTCTGATGGACAAGGACATCATGATCCTCC
ATCAACTGCTCCTTCTCTTTACATCATCATCACACAACCTCAGAAATTTGGCGATGAGACC
TCCAACATCGACGCTCAACATCTTCCATCTCAGCCTATGCACATAGAGCCACCTCCTTC
TTCTACACACAATACCGATAATACAAGATTAGTTCGGGCTGCTCAACCTAGTGTTCCAC
TCGACCAGCTTCTGACCCGTCATGGACTTGACCAATCATTCTCAGTTTCATCAACCTCC
TCAAGGTTCTAAATCCATCAAGAAGGAAGGGAACCGCAAGGGTCTTGCCCTCATCGGACCA
TGACATACCTAAATCGTCAGACCCTAAACATTGAGAAGACTAGCACAAAACAGAGAAGC
AGCAAGAAAAAGCAGATTACGTAAAAAGGCTTATGTTTCAAGCACTCGAGTCATGTAGGAT
CAAACCTGACCCAACTAGAACAAAGAGATTCAACGGGCCAGATCCCAAGGCGTATTCTTTGG
AGGGTCTCTTATAGGAGGAGATCAACAGCAAGGTGGACTACCCATTGGCCCTGGCAACAT
CAGCTCTGAAGCAGCGGTGTTGATATGGAATATGCGAGGTGGCTGGAGGAGCAGCAGAG
GCTATTAAACGAACTAAGGGTGGCAACACAAGAACACTTGTCCGAGAACGAGCTTAGGAT
GTTTGTGGACACATGTTTAGCTCATTATGACCATTTGATTAACTCAAGGCTATGGTCGC
TAAGACCGATGCTTCCACCTCATTTCTGGAGCATGGAAAACCTCAGCTGAACGTTGCTT
CTTGTGGATGGGTGSTTTCCGTCCATCGGAGATCATTAAGGTGATTGTGAACCAGATAGA
ACCATTGACGGAGCAACAGATAGTTGGGATATGTGGGCTGCAACAGTCCACACAAGAGGC
CGAGGAGGCTCTCTCGCAAGGCCTCGAGGCGTTGAATCAATCACTTTCCGATAGCATTGT
CTCTGACTCCCTCCCGCTGCTCCGACCACTTCTCTCTCATCTATCCAATTTTATGTC
ACACATGTCCTTAGCTCTCAACAAGCTCTCTGCTCTCGAGGGCTTCGTTCTCCAGGCGGA
TAATTTGAGGCACCAACGATCCATAGGCTGAACCAATTGTTGACGACCCGTCAAGAAGC
ACGGTGTCTTCTAGCCGTTGCGGAGTACTTCCACCGTCTTCAAGCTCTAAGTTCTCTCTG
GCTAGCCCGTCTCTCGGCAAGATGGATAATACTAAAACAACCTGATGAAGGAAACCAAAAAC
AAAAACAAGAGAATAGGTTGATTAGTTAGCCGCCAGCTTGACCTCTTTATCATATATATC
GTCTCTCTACTCAAATACAGTGCAATTAGGGAATTTGTTGGCTTCTTTTGGTATATG

151

GAGAAACACGACACGAATTTTGTATAATTAAGATTACACAAAAAAGTGTAGAAAGAG
AAATATCTTCTTCTTTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTTATCCTTTTTGTTC
TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAAGAACCATCGAAGTTAAATCATC
CACCGGAGTTTACGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
TAGCGGTGGCGTTGGATTTAGTCTGGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
TCTGTAGAGTTCAAGTCTTTCTCTCAGCTTTTAGCTGGAGCTATGGCTTCTCCGGCGGC
AGCTGTCTGTGCGCGCTGTGTGGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTAAGCAGAGTAGACC
AACGGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCGGGGTTAAGTCC
GGCTACTCTTTTGGATTCTCCGAGCTTCTTTGGTCTTTTTTTCACCTCTTCAGGGAACATT
TGGTATGACACATCAACAAGCTTTAGCACAAAGTCACTGCACAAGCAGTTCAAGGCAATAA
TGTTTCATATGCAGCAATCACAACAATCTGAATATCCTTCTTCTACACAACAACAACA
ACAACAACAACAAGCTTCAATTGACTGAGATTCCATCATTCTTCTGACCTAGGTCTCA
GATTCGAGCCTCGGTTCAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAATATCTGT
CTTTGAGCATCGGTACAGCCTCAAATGCTGACAAACCAGCTGATGATGGATACAACTG
GCGGAAATATGGGCGAAGCAAGTGAAGGGGAGCGATTTTCTCGGAGTTATTACAAATG
TACGCATCCAGCTTGTCTGTCAAGAAGAAAGTGGAGAGGTCACTCGATGGACAAGTAAC
GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTCAAAGCGCGGTAAACAATAA
CGGGAGTTGTAAAGTTCTGATATTGCAAAATCAGTTTCAAACAAGTAATAGCAGTCTCAA
CAAGAGTAAGAGGGGACCAGGAAACAAGTCAACAACAGAGCAGATGTCGAAGC
AAGTGATAGCGAGGAGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
GCCTGATCCCAAGCGAAGAAATACAGAAGTTGCGGTTTCAAGAACAGTTGCTTCATCGCA
TAGAATGTGACAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
TGATGGATATAGGTGGCGCAAGTATGGTCAGAAAGTAGTCAAAGGAAATCCTTATCCGAG
GAGCTACTATAAGTGTAACAACACAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAACATAACCATGATGTTCCAGC
TGCTAGAACCAGCAGCCATCAGTTAAGACCAACAATCAACACAACACCTCAACGGTTAA
CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGGCTTAAAGAAGAGCAAAATCACTTGACA
GAGAAGAAGAATACGACGGCGCTTGAGCTTTTGAGTGTAAATGAATCTTCTTTTGGTT
AATGAACCTGTTTTTGTGCTCAAAACACACAGGTTTCTCTGGACAGAATCTCTGATA
TTACAGTTTCAAAGGTATGTTCTTTTATTTCATGTTGGAATCTTCTGTGAATCTTAAG
AAGCTTTAGGAGGTAAATGTAACCAACAGATTCAAAGTTATGCCCTTATGTGAATCTTT
GTACATGGGATAAACAAATTTACAGGTATCCTTTTTTGTCTTGTGTAAAAA
AAAA
>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)
MAEKEEKEPSKLSSTGVSRPTISLPPRPFGEFFSGGVGFSPGPMTLVSNLFSDFDEFK
SFSQLLAGAMASPAAAAVAAAAVVATAHHQTPVSSVGDGGGSGGDVDPFRKQSRPTGLMI
TQPPGMFTVPPGLSPATLLDPSFFGLFSPLOQTFGMTHQQALAQVTAQAVQGNVHMQQ
SQQSEYPSSTQQQQQQQQQASLLEIPSFSSAPRSQIRASVQETSQGQRETSEISVFEHRS
QPQNADKPADDGYNWRKYQKQVKGSDFPSYKCTHPACPVKKKVERSLEDGQVTEIYK
GQHNHELPOKRGNNNGSCKSSDIANQFQTSNSSLNKSQRDQETSQVTTTEQMSEASDSEE
VGNAETSVGERHEDEPDPKRRNTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
RKYGQKVVKGNPYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
HQLRPNNQHNTSTVFNHQPPVARLRLKEEQIT*
>G1134 (61..849)
TAAAGAAAGAGAAAAAAGCTTTCGTAGTGTCTATTGAAACCAGAGAAAAAGCCAAAGGGG
ATGCAACCAACATCCGTCGGTAGTAGCGGCGGTGTGACGACGAGGAGGCAGAGGAGGA
GGAGGAGGGCTAAGTAGAAGTGGACTATCTCGGATCCGTTCACTCCAGCGACTTGGCTT
GAAGCTTTACTTGAGGAAGATGAAGAAGAGTCTTTGAAACCTAATCTTGGTCTCACCAGT
TTGCTTACCGGGAACCTCGAACGATTTACCGACAAGTCCGCGCTCGTTGAGTTCCCGATT
CCTGTTGAGCAAGGGTTGTATCAACAAGGTGGGTTTCAACGACAGAATAGTACTCCGGCG
GATTTTCTTAGTGGTTCTGATGGATTTATCCAAAGCTTTGGGATTCAGGCGAATTACGAT
TACTTATCGGGGAATATCGATGTTTCTCCGGGAAGTAAGCGGTCTAGAGAAATGGAAGCA
CTCTTCTCTTCTCTGAGTTTACTTCTCAAATGAAAGGAGAGCAAAGCAGCGGTCAAGTT
CCTACCGGAGTATCAAGCATGTGCGATATGAACATGGAGAACCCTTATGGAGGACTCTGTT
GCTTTTAGGGTTCGGGCTAAACGTGGTTGCGCAACTCATCCCGCAGCATTGCCGAGAGG
GTACGAAGGACGCGGATTAGTGATCGGATAAGGAAGCTACAAGAGCTTGACCTAACATG

GACAAGCAAACCAACTGCAGACATGTTAGAAGAAGCAGTAGAATACGTGAAAGTTCTT
CAAAGGCAGATCCAGGAGTTAACAGAAGAACAGAAGAGGTGCACATGCATACCTAAGGAA
GAACAATAAGGTTTGCTCCTGATTTGTTTTATATTTGCTTAACGGCAATGATCTGATCGA
AAAATTCGAAAGATGATCTTAGCTTGAATTTAGATGGATGTCATGTTGAAAAGTATATTA
TTTGATAAATGGATGTAGGTGTAATATAAAATTTTGTACAATAATGAAGAAAGTTAAAA
AGAATTAATGAAAACATATATCTTTATGATATAAAAAAAAAAAAA

>G1134 Amino Acid Sequence (domain in AA coordinates: 198-247)

MQPTSVGSSGGGDDGGGRGGGGLSRGLSRIRAPATWLEALLEEDEEESLKPNLGLTD
LLTGNSNDLPTSRGSFEFPIVEQGLYQGGFHRQNSTPADFLSGSDGFIQSFQIANYD
YLSGNIDVSPGSKRSREMEALFSSPEFTSQMKGEQSSQVPTGVSSMSDMNMENLMEDSV
AFRVRAKRGKATHPRISIAERVRRTRISDRIRKLQELVPNMDKQNTADMLEEAVEYVKVL
QRQIQELTEEQKRCTCIPKEEQ*

>G1008 (89..973)

GCCTTTTTGACTCTTCTTCTCTCTCTACTTTTTTTTCAGGCTCTCTCTCTATATCTCTA
TCTTCTTCTCCGGTTAACTAAAAAGAGAAATGAAAAGCCGAGTGAGAAAATCCAAGTACAC
GGTTCACCGGAAAATCACATCCACACCGTTTCGACGGTTTCCCGAAGATTGTCAAAATCAT
AGTCACTGACCCATGCGCTACTGATTCTTCCAGCGATGAGGAAAACGACAACAAATCTGT
TGCTCCGAGGGTGAAAACGTTATGTGGATGAGATCAGGTTCTGTGACGAAGATGACGAACC
TAAACCGGCGAGGAAAACGAAGAAAAAGTCCCCGGCGGCTGCGCGGAGAACGGTGGAGA
TTTGGTAAAGTCTGTGGTGAAGTATAGAGGAGTGAGACAACGACCTTGGGGAAAAATTGCG
GGCGGAGATTTCGTGATCCTTCGAGTCGTACTAGACTCTGGCTTGGGACTTTTGGCGACGGC
GGAGGAAGCTGCTATAGGTTACGATAGAGCCGCGATTTCGAATCAAAGGTCATAACGCTCA
GACGAATTTTCTCACTCCTCCTCTAGTCCGACGACTGAGGTGTTACCGGAAACTCCCGT
TCTCCGACTTCTGTCTCTGGTTGTGATTTCGGCGAGGGAATCGCAAATCAGTCTGTG
AACGGAAGAACAAAATCCGTTTTTCTTGCCTGATTTGTTTCGCTCCGAGATTATTTTTG
GGATTCCGAAATTACCCCTGACCTTTGTTTCTCGACGAATTCCACCAGTCCTTGTACC
AAACATCAACAACAACAACACAGTGTGTGATAAGGATACGAATCTGTCTGATAGTTTTCC
GTTGGGAGTGATCGGAGATTTCAGCTCATGGGATGTTGATGAGTTTTTCCAAGATCATT
GTTGGATAAGTAATTTGATGAGTTCTTCCCCAGAATTTTTCTGGGTTTCTCTTTTGGTT
GTGTGAGTGAGATGAGTGGTTTGTATGACAACGACGGGGATGAATCTTAGCCGTCGGTTTT
CCATTTCCGTGGACGGCTCCGATCAGCGGAAGAAGCGCAACGGAGTTTTTATTTATCTGTT
TGAGAATTTTATAATTTAATTTGCGAGTAAATATAGTAATTAGTGTTAAGATTGTGAGAG
TTTAAGTTAATTAGGGAGGGGTTTTGAATATTGGGGATTTTGGGAGGTTTTGTTTGGTT
TCTCTCCAAGTCTGTCACTATGCAAGGAAGCAGTATAAAGACCGTATATATATTTTATTA
TTAATATTGATAAAAGTAAAAAAAAAAAAAAAAAAAA

>G1008 Amino Acid Sequence (domain in AA coordinates: 96-163)

MKSRVRKSKYTIVHRKITSTPFDGFPKIVKIIIVTDPCATDSSSDEENDNKSVAAPRVKRYVD
EIRFCDEDEPKPARKAKKKSAAAAENGDDLKSVVKYRGVRQRPWGKFAABIRDPSSR
TRLWLGTFAATAEEAIGYDRAAIRIKGHNAQTNFLTTPPPSPTEVLPEPVIDLETVSGC
DSARESQISLCSPTSVLRFSHNDETEYRTEPTTEQNPFPLPDLFRSGDYFDSEITPDPL
FLDEFHQSLLPNNNNNTVCDKDTNLSDSFPLGVIGDFSSWDVDEFFQDHLDDK*

>G1020 (132..689)

CTGTTTACAAGAAAGCTCCCCAAAAGGAGCGTTGCTTTACTCTCTCTATAAAAAGAGCTC
TTCTACTTCTTCTCGTTACCACAAAACCTCTTTACCGATCTTCTCGTTCCATTCTTCTTC
CTAATTACACCATGCCCAACATCACCATGGGTTTGAAACCCGACCCGGTTGCTCCAACGA
ACCCGACTCATCATGAGAGTAATGCTGCCAAAGAGATTCTGTACAGAGGCGTTAGGAAAC
GTCCATGGGGAAGATACGCCGCTGAGATCCGAGATCCGGTTAAGAAAACCTGAGTCTGGC
TCGGTACGTTTCGACACCGCTCAGCAGGCGCGCGTGCTTACGACGCGAGCCGCGCGTGACT
TTCGTGGTGTTAAGGCTAAGACCAATTTTCGGTGTTATCGTTGGTAGTAGTCTACTCAGA
GTAGCACCGTCGTCGACTCTCCACGGCGGCACGGTTTTATAACACCTCCGCACCTCGAGC
TCAGCTTAGGCGGCGCGCGCGTGTCGTGTAAGATCCCGCTTGTGCATCCGGTTTACT
ACTATAACATGGCGACGTATCCAAAGATGACGACGTGTGGTGTCCAGAGCGAGTCTGAAA
CGTCGTGCGTCGTTGATTTGGAAGGTGGAGCTGGGAAGATATCTCCGCCGTTAGATCTGG
ATCTTAACCTTAGCTCCTCCGGCGGAATAGGCCGTGAGTTTTTTTTTTCTTATGTCTGTTT
TTTAGACAAAAAAAATAACGTTTCTTTTTTTTCTGCCTAAGAAAAAATATTATCCG
TTTTTTAGAAGAAAAAAAATAAAAAAAAAAAAA

>G1020 Amino Acid Sequence (domain in AA coordinates:28-95)
 MPNITMGLKPDVPATNPHTHESNAAKEIRYRGVRKRPWGRYAAEIRDPVKKTRVWLGT
 DTAQQAARAYDAAARDFRGVKAKTNFGVIVGSSPTQSSTVVDSPTAARFITPPHLELSLG
 GGGACRRKIPLVHPVYYNMATYPKMTTCGVQSESETSSVVDFFEGGAGKISPPLDLNL
 APPAE*

>G1023 (252..1250)
 TCGTCTTCTTAATCGCTTTCTGCTCTGTTTTCTCGTTCATCAAGCTACATCTACTAGCT
 CTCTCAGTGATTGATTTCTCACAGTTTCATCGATTTCCATGCGTTTAAGACCTAAAAGGA
 CTTGTCTCGGGTAAAGGACTTTTCTGTTCTTGAGAGAGTTCATTTTGAGGCTTTTCTG
 GGAATTTTGAGAGGTTTTTGGGTTTAAGGGGGTTTGGTTTGAATTTTCGCACACCAAG
 TGTTCGATAAAATGGCTGAACGAAAGAAACGCTCTTCTATTCAAACCAATAAACCAACA
 AAAAACCCTGAAGAAGAAACCTTTTCAGCTAAATCACCTCCCAGGTTTATCTGAAGATT
 TGAAGACTATGAGAAAACCTCCGTTTCGTTGTGAATGATCCTTACGCTACTGACTACTCAT
 CAAGCGAAGAAGAAGAAAGGAGTCAGAGAAGGAAACGTTATGTCTGTGAGATCGATCTTC
 CTTTCGCTCAAGCTGCTACTCAAGCAGAATCTGAAAGCTCATATTGTGAGGAGAGTAACA
 ATAATGGTGTAAAGCAAGACTAAAATCTCAGCTTGTAGCAAAAAGGTTTTACGCAGCAAAG
 CATCTCCGGTCGTTGGACGTTCTTCTACTACTGTCTCGAAGCCTGTTGGTGTAGGCAGA
 GGAAATGGGGTAAATGGGCTGCTGAGATTAGACATCCAATCACCAGTAAGAATTTGGT
 TGGGTACTTACGAGACGCTTGAACAAGCAGCTGATGCTTATGTACCAAGAAGCTTGAGT
 TTGATGCTCTGCTGCGAGCCACTTCTGCTGCTTCTCTGTTTGTCAAATGAGTCTGGTT
 CTATGATCTCAGCCTCAGGGTCAAGCATTGATCTTGACAAGAAGCTAGTTGATTGACTC
 TTGATCAACAAGCTGGTGAATCGAAGAAAGCGAGTTTTGATTTGACTTTGCGAGATCTAC
 AGATTCCTGAAATGGGTGCTTATTGATGACTCATTATCCCAAATGCTTGTGAGCTTG
 ATTTTCTCTTAACAGAAGAGAACAAACCAAAATGTTGGATGATTACTGTGGCATAGATG
 ATCTGGACATCATTTGGTCTTGAATGTGACGGTCCAAGCGAACTTCCAGACTATGATTTCT
 CAGATGTGGAGATCGATCTTGGTCTCATTTGAACCACTTGAAGATGCTTTTCGTTG
 ATCATATCGCAACAACCTACTCCCACTCCTTAAATATCGCGTGCCCATAAAGTTTTCAGC
 TAGGTGTTATTATTAGCTATAGGAGCAACGTAAAAAGCTCGTTGTTACTCGGTTTTGTCT
 TAAGTTATTAAAGTATAGCAGAGGCAGTTAATCTCAAGGGAAGCAAAAACCTAAAGATA
 GAAGCAGATGCAGTTTTGTGTGTTGGTGTACTAAAGAAAGTTTTGTTGACATAATGGTT
 TTGATGTTGTGGAGAAGATAGAGAGGTGTGATCGAAATTGTAAATCTCAGGTGGTTTTTT
 TTGAAGGCAATTGTTTCTATTAGGGTTTTTTTCTATATGAGGATTGTCTTTGAAAAGC
 CTTTAGATGTTTTCTAATTCGTAAGCTCTCTCAATCTTTGTAAGTTTTGCCTGTTGAGTT
 ATTGATACATATGTGAGACCTACTTTATTGTTTTGTGCTACATACATTGTTGATGGTTT
 CGTCAAAAAAAA

>G1023 Amino Acid Sequence (conserved domain in AA coordinates:128-195)
 MAERKKRSSIQTNKPNKPKMKPFQLNHLPLGSEDLTMRKLRVNDPYATDYSSSEE
 EERSQRRKRYVCEIDLPPQAATQAESESSYQESNNNGVSKTKISACSKVLRSKASPV
 VGRSSTTVSKPVGVQRKWKWAAEIRHPITKVRTWLGTYETLEQAADAYATKKLEFDAL
 AAATSAASSVLSNESGSMISASGSSIDLKLVDSLTLDQQAGESKKASFDFDFADLQIPE
 MGCFFIDDSFIPNACELDFLLTEENNNQMLDDYCGIDDLDIIGLECDGPSELDPDYFSDVE
 IDLGLIGTTIDKYAFVDHIATTTPTPLNIACP*

>G1053 (38..538)
 GAAACTCTTACATACTCATATAAACCAAACTAAAACCATGATTCCGGCAGAAATCAACGG
 ATATTTCCAATATCTATACCCGAATACAACGTAATAAACATGCCTTCATCTCCAACCTC
 TTCTTTAACTACCTAAACGATTGATCATCAACAACAACAACTATTCTCATCATCCAA
 CAGTCAAGATCTCATGATAAGCAACAACCTCAACTTCCGACGAAGATCATCATCAAAGCAT
 CATGGTACTCGACGAGAGGAAACAGAGAAGGATGCTTTGAAACAGAGAATCTGCAAGGAG
 GTCAAGGATGAGGAAACAGAGACATCTTGATGAACCTCTGGTCTCAGGTAATAAGGCTTCG
 CAACGAGAACAACCTGTCTTATCGATAAGCTGAACCGCGTATCGGAGACTCAAAATTGTGT
 ATTGAAGGAGAACTCTAACTCAAGAAGAAGCTTCTGATCTCCGACAGCTTGTGTGTA
 ACTGAAATCTAACAAGAACAACAATAAGTTTTCCAAGAGAGTTTGAAGATAATTAGTA
 TTACTCAA

>G1053 Amino Acid Sequence (domain in AA coordinates: 74-120)
 MIPAEINGYFYLSPEYNVINMPSSPTSSLNYLNDLIINNYYSSSSNSQDLMISNNSTS
 DEDHHQSIMVLDERKQRRMLSNRESARRSRMRKQRLHDELWSQVIRLRNENNNCLIDKLN
 VSETQNCVLKENSCLKKEASDLRQLVCELKSNKNNNNNSFPREFEDN*

>G1137 (202..1248)

TACTTCAGACTTCTACTCAAACCACTCACGTAAGTGGTTGGTGACATTTTCGCTGCATTTT
TCAATCTGTGATTGTTTTTCGTTTCGTCCTTTCTTTACTATTTTCTCGAAAAGGACACAAG
AAGTATTGCATTCACTCAGTTGAGCAACTTAACAATCGTGTGTACTTTTTGAAGTTCCTC
TTGAGCTAAACTGCTAAGAGCATGCCTCTGGATAAGAGGCAACGGGATTTGCCTCTGGGC
TTAAGTCTCAAGCTTGCTTCAAGGATATAGTAGGTCGGTCTGTCCTTCCTAGAATTCCT
CTCCCTGAGCTTGGGAACTATATGCAAGCTAAGCTTCAGGCTCGCTGTTTGCAGCCACCA
CCATTCCAGTCTTTTGCTGTGCAGTCATGATAAGGAGTCTTATGGAAAAAGATTCTCACGG
TCTGACATGCGGTCTTGGTGCCTGCTGCTACTACTACTACTCCACTTGGAGCATT
GAGTCTTCTCAGAAAAGACTTTTGATATTCGATCAGTCAGGAGACCAGACTCGTCTATTA
CAATGTCCATTTCTCTACGGTTTCCATCTCATGCGGCTGCAGAACCACTGAAACTCTCT
GAGTTACAAGGTATAGAGAAAGCTTTCAAAGAAGATGGTGAAGAGTTTCAAGAGTGAT
GGAACAGAGTCAGAAATGCATGAAGACACTGAGGAGATCAATGCATTGCTATATTAGAT
GATGATTATGATGATGATTGCGAGAGTGATGATGAAGTAATGAGCACTGGTCACTCTCCT
TATCCAAATGAAGGAGTTTGCAACAAAAGGGAATTAGAAGAAATCGATGGTCTCTGTAAA
AGGCAGAACTACTGGATAAGGTCAACAACATCAGCGACTTATCATCACTTGTGGGCACT
GAGAGCTCCACACAACCTCAATGGATCTTCTTTCTTAAGGACAAAAGCTCCCTGAATCA
AAAACCATATCGACCAAGAGGACACTGGTCTGGTCTGAGCAACGAGCAGTCGAAGAAA
GACAAGATCCGCACAGCTCTGAAAATACTCGAGAGCGTAGTCCCTGGTGCAAAAAGGAAAC
GAAGCGCTCTTACTTCTGGACGAAGCAATTGATTACCTAAAGTTGCTGAAACGAGACTTA
ATCTCCACAGAGGTTAAGAACCAGCTCCACCACTCACAAGTCACCAATCTTGTGCTT
AAAGAGACAACATGGGGAACAGAAATCTGCAGACAGATAAGGCGTGAAAGATTCTGACG
AGTTAAACCGTGTGAAGTGGGTTTTTGGGTACGTATCCTTGCACCAGCTTT

>G1137 Amino Acid Sequence (domain in AA coordinates: 264-314)
MPLDKRQRLPLGLSPQACFKDIVGRSVLPRIPLPELGKLYAAKLQARCLQPPPFQSLLC
SHDKESYGRFSSRSDMRSWCAAAATTTTTPLGALESSQKRLLIQDQSGDQTRLLQCPFPLR
FPSHAAEPVKLSELQIEKAFKEDGEEFHKSDGTESEMHEDTEEINALLYSDDDYDDDC
ESDDEVMSTGHSPYPNEGVCNKRELEEIDGPCKRQKLLDKVNNISDLSSLVGTESSTQLN
GSSFLKDKKLPESKTISTKEDTSGSLSNEQSKDKIRTKILESVPVPGAKGNEALLLLD
EADYLLKRLDLISTEVKNQSSSTHKSPILLKETTGWTRNLQTDKA*

>G1181 (113..1012)

CTCGATCTTTTAACCCCATTTATTACATATTACTCCTTCCTACATTATTCTTCTCTGCT
TTCGTGACTTTTACGGGGACACTTTTGTTTTATAACTTACGCTTAAATCCTATGAATTC
GCCGCCGTTGACGCAATGATTACCGGAGAATCATCGTCACAAAGATCTATCCCAACGCC
GTTTCTCACAAAAACGTTTAACTCGTTGAAGATAGTTCCATCGACGATGTTATCTCATG
GAACGAAGATGGTTCCTCTTTCATCGTATGGAATCCGACAGATTTGCTAAAGATTTGCT
TCCTAAACACTTCAAACACAACAATTTCTCTAGTTTCGTTTCGTCAGCTCAACACTTACGG
ATTCAAAAAAGTTGTACCGGATCGATGGGAGTTTTCAACGATTTCTTTAAGAGAGGAGA
AAAACGCTCTTCCCGTGAGATCCAACGTCGGAATAACAACGACGCATCAAACAGTTGT
TGCTCCTTCGTCGGAACAACGAAACGACGATGGTTGTATCACCGTCAAATTCGGGGGA
AGATAATAATAAATCAGGTGATGTCTTCGTCCTCGTCGTCGTTGTTATGTCATCAAAC
GAAGACGACTGGGAATGGTGGTTTATCAGTGGAGTTATTGGAAGAGAACGAGAAGCTTCG
GAGTCAAAACATTCAGCTAAACCGTGAGCTTACTCAGATGAAATCTATCTGCGATAATAT
CTATAGTCTCATGTGCAATTACGTGCGATCTCAGCCCACTGATCGGAGTTATTCTCCCGG
AGGTAGTAGTAGTCAACCGATGGAGTTTTTACCGGCGAAGCGGTTTTTCGGAGATGGAGAT
TGAAGAAGAAGAAGAAGCGAGTCCGAGGTGTTTGGTGTTCGATTGGGTTAAAACGGAC
GAGAAGTGAAGGTGTTCAAGTGAAGACGACGCGGTGGTTGGGGAAAATTCGGATGAGGA
GACGCCGTGGTTGAGACATTATAATCGAACCAATCAGAGAGTTTGTAAATAAAAACGAAC
GGTTTAGATTTGTGGTGTAGATATGTGCGCGAAGTAGACGATTACAGCTTTTTAAGACAA
GCAGAGACGTTGTCCTCTGTTTCAAGAAGTTTCTGCAATCTTGACTTCTTCTTTTAAC
ACTTGTGTTTTTTTATTATTAAATTAATAAATAAATGTTCTTTTTCAGTTTTGTTTTC
TTCAAAAATAGTTGCGCTGTTTCTAGACTTTCTTTTTTT

>G1181 Amino Acid Sequence (domain in AA coordinates: 24-114)
MNSPPVDAMITGESSQRSIPTPLTKTFNLVEDSSIDDVISWNEDGSSFIWNPDTFAK
DLLPKHFKHNNFSSFVRQLNTYGFKKVVPDRWEFSNDFFKRGEKRLREIQRKITTTHQ
TVVAPSSEQRNQTMVVSNSNGEDNNNNQVMSSSPSSWYCHQTKTTGNGLSVELLENE
KLRSQNIQLNRELQMKSI CDNIYSLMSNYVGSQPTDRSYSPGSSSQPMEFLPAKRFSE

MEIEEEEEASPRFLGVPIGLKRTRSEGVQVKTTAVVGENSDEETPWLRHYNRTNQRVCN*
>G1228 (63..1139)
GCATTATAATTACTCACTCATCTCTCTTTTCATTACATTACATACCAAACAAGAGCTCTC
AAATGGAAAGGTTTCAAGGACACATCAACCCCTGTTTCTTCGATCGAAAACCGGATGTGA
GAAGCCTCGAGGTTCAAGGATTTGCAGAGGCTCAAAGCTTTGCTTTCAAAGAAAAAGAGG
AAGAAAGCTTACAAGATACAGTTCCATTTCTACAGATGCTGCAAAGTGAAGACCCCTCAT
CGTTTTTTTCAATCAAAGAGCCAAACTTTCTGACGCTACTGTCTCTTCAAACCCCTCAAGG
AGCCTTGGGAACTCGAAAGATATCTTTCACTTGAGGATTACAATTTCAATTCACCGGTCC
AATCTGAGACCAACCGCTTCATGGAAGGAGCCAATCAAGCTGTGTCAAGCCAAGAAATTC
CCTTTAGCCAAGCAAACATGACACTCCCTTCTTCTACCTCATCACCCTCAGTGCACATT
CAAGACGAAAGCGCAAAATCAACCCTTGCTGCCTCAAGAAATGACTAGAGAAAAGAGAA
AGAGGAGGAAAAACAAACCAAGTAAAAACAATGAAGAGATTGAGAATCAAAGATAAACC
ACATTGCTGTTGAACGAAACAGAAGACGTCAAATGAACGAACATATCAACTCTCTCCGGG
CCCTTCTCCCACTTCTTACATCCAACGAGGAGACCAAGCTTCCATAGTAGGAGGAGCAA
TAACTACGTGAAGGTCCTCGAGCAAATCATACAATCTCTCGAATCGCAAAGAGAACGC
AACAACAAAGTAACAGTGAGGTAGTAGAAAACGCCTTAATCATCTCTCAGGCATTTTCGT
CGAACGACCTGTGGACAACCTTTGAAGATCAAACCTTGATCCCCAAAATCGAAGCTACAG
TGATACAAAACCATGTGAGCCTTAAAGTTCAATGTGAGAAGAAACAAGGACAACCTCTCA
AAGGAATCATATCACTTGAAAAGCTTAAACTCACTGTTCTTCATCTCAATATCACTACTT
CGTCTCATTCCTCTGTTTCTTATTCTTCAACCTCAAGATGGAAGATGAGTGCAGCTTAG
AGTCAGCCGACGAGATTACGGCGGCTGTTTCATCGGATTTTCGATATTCGACAATTTGAT
TAAACACATATAATTCCAAAAATATTAACAGCTGACAAAATGGTATCTTTGCGGCC
>G1228 Amino Acid Sequence (domain in AA coordinates: 179-233)
MERFQGHINPCFFDRKPDVRSLEVQGFAEAQSFQFKEKEEESLQDTPVFLQMLQSEDPSS
FFSLIEPNFLTLTSLQLKEPWELERYLSLEDSQFHSPVQSETNRFMEGANQAVSSQEIP
FSQANMTLPSSTSSPLSAHSRRKRKINHLLPQEMTREKRKRRTKPSKNNEEIEENQRINH
IAVERNRRRQMNHINSLRALLPPSYIQRGDQASIVGGAINYVKVLEQLIQSLESQKRTQ
QQNSEVVENALNHLSSISNDLWTTLEDQTCIPKIEATVIQNHVSLKVQCEKKQGGQLLK
GIISLEKLLTFLHLNLTSSSHSSVSYSFNLKMEDECDLESADETAAVHRIFDIPTI*
>G1277 (51..512)
ATTCTAAAGTCTCTCTCGGAAAGTAAGAGACTCAACTTCCGAGCCGCCATGACGCGG
GAGTAGCAGTAAAGCTGACGTGGCAGTCAAAATGAAGAGAGAAAGACCATTCAAAGGGA
TCAGAATGAGAAAATGGGGGAAATGGGTTGCGGAGATTGAGAACCCCAACAGCGTTCAA
GACTTTGGCTCGGCTCTTACTCTACTCCCAGCGGCGGCGCGTGCATACGACACGGCTG
TCTTTTACCTCAGAGGACCAACTGCTACGCTCAACTTCCCGGAGCTTCTGCCGTGTACCT
CCGCCGAGGATATGTGACGCGCAACGATCAGGAAAAAGGCGACGGAGGTGGGAGCTCAAG
TAGATGCGATAGGGGCGACGGTGGTGCAGAACAAACACGCCCGCGCGTCTTTTAGTCAAA
AGCGTGACTTTGGCGGCGGGTTATTAGAGCTTGTGACTTGAACAAGTTACCTGACCCGG
AAAATCTCGATGATGATTTGGTGGGAAAATAGACTGAAAAATAATAATAAATATCTTAC
AATGGTGGCTGTAGCTATCGTACGCGGAATGCTTGGGCTTGTGTTATATGACTACGTGGT
TACGGAAGATTCTCTGTTTCGTCATTGTATTAAAAATTAATCCCAAGTCAAACATA
CTGTACATTATTCTTAATTTAGTATTTTCTTATTAATATCTATCATTTGTTTGGTGAACA
CCAGAATATTAGACTATTAATGTAAACGAGTTTTTAATATTTGATCATAATAACACCAAG
CTAGTTAAAGGTTAATATCTTGTACGAAGTCTTGAGTAAGTTCAATTGTATATATATG
TAACGGAAGAGGTTCTGTTCCGGTCCCAAGTGAAGTGGATCAAAGGTGACTTCACATAAAA
AATAAAAAAAA
>G1277 Amino Acid Sequence (domain in AA coordinates: 18-85)
MDAGVAVKADVAVKMKRERPFKGIRMRKWGWVABIREPNKRSRLWLGSYSTPEAAARAY
DTAVFYLRGPTATLNFPELLPCTSAEDMSAATIRKKATEVGAQVDAIGATVVQNNKRRRV
FSQKRDFGGGLLELVDLNKLDPDENLDDDLVGK*
>G1309 (53..859)
CGTCGACCTCTTAATTAAGACGACTTGAGAGAGAAAGAAAGATACGTGGAAGATGACCAA
ATCTGGAGAGAGACCAAAACAGAGACAGAGGAAAGGTTATGGTCACCTGAAGAAGACCA
GAAGCTCAAGAGTTTCATCTCTCTCGTGGCCATGCTTGCTGGACCACTGTTCCCATCTT
AGCTGGATTGCAAAGGAATGGGAAAAGCTGCAGATTAAAGGTGGATTAATTACCTAAGACC
AGGACTAAAGAGGGGGTCTGTTTAGTGAAGAAGAAGAAGAGACCATCTTGACTTTACATT
TTCTTGGGTAACAAGTGGTCTCGGATTGCAAAATATTACCGGGAAGAACAGACAACGA

GATTAAGAACTATTGGCATTCTATCTGAAGAAGAGATGGCTCAAATCTCAACCACAACT
 CAAAAGCCAAATATCAGACCTCACAGAATCTCCTTCTTCACTACTTTCTTGCGGGAAAAG
 AAATCTGGAAACCGAAACCCTAGATCACGTGATCTCCTTCCAGAAATTTTCAGAGAATCC
 AACTTCATCACCATCCAAAGAAAGCAACAACAACATGATCATGAACAACAGTAATAAAGT
 GCCTAAACTGTTCTTCTCTGAGTGGATCAGTTCTTCAAATCCACACATCGATTACTCCTC
 TGCTTTTACAGATTCCAAGCACATTAATGAACTCAAGATCAAATCAATGAAGAGGAAGT
 GATGATGATCAATAACAACAACACTCTTCACTTGAGGATGTCATGCTCCGTACAGATTT
 TTTGCAGCCTGATCATGAATATGCAAATTATTATTCTTCTGGAGATTTCTTCATCAACAG
 TGACCAAAATTATGTCTAAGAAGAGTGAATATGATCGTAAGAGGAACATAAGCTAGTTAC
 TTGTGTTACAGC

>G1309 Amino Acid Sequence (domain in AA coordinates: 9-114)
 MTKSGERPKQRQRKGLWSPEEDQKLKSFILSRGHACWTTVPILAGLQRNGKSCRLRWINY
 LRPGLKRGSFSEEEEEITLTLHSSLGNKWSRIAKYLPGRDNEIKNYWHSYLKKRWLKSQ
 PQLKSQISDLTESPSSLLSCGKRNLLETETLDHVISFQKFSENPTSSPSKESNNNMIMNNS
 NNLPLKFFSEWISSNPHIDYSSAFTDSKHINETQDQINEEBVMNNNNYSSLEDVMLR
 TDFLQPDHEYANYYSSGDFFINSDQNYV*

>G1314 (1..990)

ATGGGAAGAGCTCCGTGTTGCGACAAGACAAAAGTGAAGCGAGGGCCTTGCTCGCCTGAA
 GAAGACTCTAACTTAGAGATTACATTGAAAAGTATGGTAATGGTGAAATTGGATCTCT
 TTCCCTCTCAAAGCCGGTTTGAGGAGATGTGGGAAGAGTTGTAGACTGAGGTGGCTAAAC
 TATTTGAGACCAAAACATAAAGCATGGTGACTTCTCTGAGGAAGAAGACAGGATCATTTTT
 AGTCTCTTCGTGCCATAGGAAGCAGGTGGTCAATAATAGCAGCTCATCTACCGGGACGA
 ACAGACAACGACATAAAAACTATTGGAACACAAAGCTAAGGAAGAACTCTTGTCTTCT
 TCCTCTGATTTCATCATCATCAGCCATGGCTTCTCTTATCTAAACCTATTTCTCAGGAT
 GTGAAAAGACCAACCTCACCAACAACAATCCCATCTTCTTCTTACAATCCGTATGCTGAA
 AACCTTAATCAATACCCAACAAAATCCCTCATCTCCAGCATCAATGGCTTCAAGCTGGT
 GACAAACAGATAATTTCTTATATTAACCTAATTTATCTCAAGATCTCTATCTCTCGGAC
 AGCAACAACAACACCTCGAACGCAAATGGTTTCTTGCTCAACCACAATATGTGTGATCAG
 TACAAGAACCACACAGTTTCTTCTCAGACGTCAATGGGATAAGATCAGAGATTATGATG
 AAGCAAGAAGAGATAATGATGATGATGATGATAGACCACCACATTGACCAGAGGACAAAA
 GGGTACAATGGGGAATTCACACAAGGGTATTATAATTACTACAATGGGCATGGGGATTG
 AAGCAAATGATTAGTGGAACAGGCACTAATTCTAACATAAACATGGGTGGTTCAGGTTCA
 TCTTCTAGTTTCGATAAGCAACCTAGCTGAGAACAAAAGCAGTGGTAGCCTCTACTAGAA
 TACAAATGCTTGCCCTATTTCTACTCCTAG

>G1314 Amino Acid Sequence (domain in AA coordinates: 14-116)
 MGRAPCCDKTKVGRPWSPEEDSKLRDYIEKYGNNGNWSIFPLKAGLRRCGKSCRLRWLN
 YLRPNIKHGDFFSEEDRIIFSLFAAIGSRWSIIAHLPGRTDNDIKNYWNTKLKRLKLLSS
 SSDSSSSAMASPYLNPISQDVKRPTSPPTIPSSSYNPAENPNQYPTKSLISSINGFEAG
 DKQIISYINPNYPQDLYLSDSNNTSNANGFLLNHNMCQYKNHTSFSSDVNGIRSEIMM
 KQEEIMMMMHIDHIDQRTKGYNGEFTQGYNYNGHGLDKQMISGTGTNSNINMGGSGS
 SSSSISNLAENKSSGSLLEKCLPYFYS*

>G1317 (1..849)

ATGGGAAGATCACCTTGTTGTGATAAAAATGGAGTGAAGAAGGGACCATGGACTGCTGAG
 GAGGATCAGAACTCATCGATTATATTCGATTTTCATGGTCTGGCAATTGGCGTACGCTC
 CCCAAAAATGCTGGACTCCATAGATGTGGAAGAGCTGCCGTCTTCGATGGACCAATTAT
 CTAAGACCGGACATCAAGAGAGGAAGATTCTCGTTTCGAGGAAGAAGAACTATCATTCAG
 CTACACAGTGTTATGGGAACAAGTGGTCAGCAATAGCCGCTCGTCTACCAGGGAGGACC
 GATAACGAAATAAAAACCATTGGAACACTCACATCCGCAAGAGACTTGTAAGGAGTGGT
 ATCGACCTGTACTACTCTCCACGCCTTGATCTTCTTGATTGTCTCTCACTTTTGAGT
 GCACTTTTCAACGAGCAAACTTTTCAGCAGTTGCAACACATGCGTCTTCTCTCTTAAT
 CCTGATGTATTGAGGTTGGCCTCTCTACTACTGCCACTTCAAAACCTAATCCAGTTTAC
 CCATCGAACCTCGACCAAAATCTTCAAACCTCAAATACATCATCAGAATCGTCTCAACCA
 CAAGCTGAGACTAGTACAGTCCCAACAACTATGAACTTCATCATTGGAGCCTATGAAC
 GCAAGACTCGACGACGTTGGTCTTGAGATGTATTACCACCTTTGTGAGAGAGTTTGAC
 TTAGACTCGTTCATGTCAACGCCAATGTCTTCTCCACGACAAAATAGCATTGAAGCAGAA
 ACCAACTCCAGCACTTTCTTCGACTTTGGAATTCCGGAAGATTTCATCTTAGATGACTTT
 ATGTTTTAA

>G1317 Amino Acid Sequence (conserved domain in AA coordinates:13-118)

MGRSPCCDKNGVKKGPWTAEEEDQKLIDYIRFHGPGNWRTPKPNAGLHRCGKSCRLRWNTY
LRPDIKRGRFSFEEEEETIIQLHSVMGNKWSAIAARLPGRTDNEIKNHNWTHIRKRLVRS
IDPVTHSPRLDLLDLSSLLSALFNQPNFSAVATHASSLLNPDVLRASLLPLQNPVY
PSNLDQNLQTPNTSSESSQPAETSTVPTNYETSSLEPMNARLDDVGLADVLPPLSESF
LDSLMTSPMSSPRQNSIEAETNSSTFFDFGIPEDFILDDFM*

>G1323 (49..870)

AAGAGGGAATCTCAAAAGTGTGTGTCTGTGAGAGAGGAGAGAGAGAATATGGGCAAAGGA
AGAGCACCATGTTGTGACAAAACCAAAGTGAAGAGAGGACCATGGAGCCATGATGAAGAC
TTGAAACTCATCTCTTTTCATTACAGAATGGTCATGAGAATTGGAGATCTCTCCCAAAG
CAAGCTGGATTGTTGAGGTGTGGCAAGAGTTGTCTGTCTGCGATGGATTAATTACCTCAGA
CCTGATGTGAAACGTGGCAATTTCAAGTGCAGAGGAAGAAGACACCATCATCAAACCTTAC
CAGAGCTTTGGTAACAAGTGGTCGAAGATTGCTTCTAAGCTGCCTGGAAGAACAGACAAT
GAGATCAAGAATGTGTGGCATAACATCTCAAGAAAAGATTGAGCTCGGAAACTAACCTT
AATGCCGATGAAGCGGGTTCAAAAGGTTCTTTGAATGAAGAAGAGAACTCTCAAGAGTCA
TCTCCAAATGCTTCAATGTCTTTTGCTGGTTCCAACATTTCAAGCAAAGACGATGATGCA
CAGATAAGTCAAATGTTTGTGAGCACATTCTAACTTATAGCGAGTTTACGGGGATGTTACAA
GAGGTAGACAAACAGAGCTGCTGGAGATGCCTTTTGTATTAGATCCTGACATTTGGAGT
TTCATAGATGGTTTCAAGACTCATTCCAACAACAGAGAAACAGAGCTCTTCAAGAGTCTGAA
GAAGATGAAGTTGATAAATGGTTTAAAGCACCTGGAAAGCGAACTCGGGTTAGAAGAAAAC
GATAACCAACAACAACAACAGCATAAACAGGGAACAGAAGATGAACATTATCATCACTC
TTGGAGAGTTACGAGCTCCTCATACATTAATGAAGCCATAAAGCAAGTCATTTTACCTT
GAAAATGGAATTTATAGCTAACTTATTGGCATTATTAGTATATAAGCAAGATCAGATAGG
CGCATGTAGTAGCAACGAAGAAACGTCGAATTGTAGACAAAATGTAGATATTACAGA
GTTGAAAGATTGTATTTTGCAAATGATTGCTTTGTAGTGAATCAAGTTATCACAAAAA
AAAAAAA

>G1323 Amino Acid Sequence (domain in AA coordinates: 15-116)

MGKGRAPCCDKTKVKRGPWSHDEDLKLISFIHKNHENWRSPLPKQAGLLRCGKSCRLRWI
NYLRPDVVRGNFSAEEEDTIIKLHQSFGNKWSKIASKLPGRTDNEIKNVWHTLKKRLSS
ETNLNADEAGSKGSLNEEENSQESSPNASMSFAGSNISSKDDDAQISQMFHILTYSEFT
GMLQEVDPKPELLEMPFDLPDIWSFIDGSDSFQQPENRALQSESEDEVDKWKHLESELG
LBENDNQQQQHKQGTEDHSSSLLSEYELLIH*

>G1332 (1..606)

ATGGAATGCAAAAGAGAAGAAGGGAAGTCTTACGTGAAGAGAGGGTTGTGGAAACCAGAA
GAAGATATGATATTAAGAAAGCTATGTTGAGACTCATGGTGAAGGAACTGGGCAGACATT
TCTCGTAGATCCGGGTTGAAGAGAGGAGGAAAAAGCTGTAGGCTGAGATGGAAGAACTAT
CTAAGACCAAATATCAAAAGAGGAAGCATGTCAACCAAGAACAAGACCTTATCATCCGC
ATGCATAAGCTTCTTGGAAACAGATGGTCGTTGATCGCTGGTCGCCTTCCAGGTCGTACT
GACAATGAAGTGAAGAACTACTGGAATACTCATTTGAACAAGAAACCTAATTCCCGAAAA
CAGAATGCACCTGAATCAATCGTCGGCGCCACTCCTTTCAACGATAAGCCAGTTATGTCT
ACAGAAGTGAAGAAGAAGCCATGGAGAAGGAGGAGAAGAGGAGCAATACCTGGATGGAG
GAGACCAACCACTTTGGCTATGACGTCCACGTAGGATCTCCCTTGCCACTTATTTCCAC
TACCCAGACAACACTCTCGTGTGTTGACCCATGTTTTCTTTTACCGATTCTTTCTCTCTG
CTTTAG

>G1332 Amino Acid Sequence (conserved domain in AA coordinates:13-116)

MECKREEGKSYVKRGLWKPEEDMILKSYVETHGEGNWADISRRSGLKRGGKSCRLRWKNY
LRPNIKRGSMSPOEQDLIIRMHKLGNRWSLIAGRLPGRTDNEVKNYWNTHLNKKPNSRK
QNAPESIVGATPFTDPKVMSTELRRSHGEGGEEESNTWMEETNHFYDVHVGSPLPLISH
YPDNTLVFDPFCFSFTDFFPLL*

>G1334 (76..885)

ATAGCTCCCAACTAATAGGAATCTCAAGCTTCTCACTCTCTCTTGTTTTCCATTGGACT
TTTGGAAACATAAGCTATGCAAACTGAGGAGCTTTTGTGCGCCACCACAGACTCCTTGGTGG
AATGCTTTTGGATCTCAGCCGTTGACTACAGAGAGCCTTTCCGGCGAAGCTTCTGATTCA
TTCACCGGAGTTAAGGCAGTTACTACGGAGGCAGAACAAAGGTGTGGTGGATAAAACAACT
TCTACAACCTCTCTTCACTTTCTACCTGGTGGTGAAGAGATTCAAGAGATGTGCCAAAG
CCTCATGTTGCTTTGCGGATGCAATCAGCTTGCTTCGAGTTTGGATTGCTCAGCCAATG
ATGTACACAAAGCATCCTCATGTTGAACAATACTATGGAGTTGTTTCAGCATACGGATCT

CAGAGGTCTTCGGGCCGAGTAATGATTCCACTGAAGATGGAGACAGAAGAAGATGGTACC
ATCTATGTGAACCTAAAGCAGTACCATGGAATTATCAGGCGACGCCAGTCCCGAGCAAAG
GCTGAAAACTGAGTAGATGCCGTAAGCCATATATGCATCACTCACGCCATCTCCATGCT
ATGCGCCGTCCTAGAGGATCTGGCGGGCGTTTCTTGAACACCAAGACAGCTGATGCGGCT
AAGCAGTCTAAGCCGAGTAATTCTCAGAGTTCTGAAGTCTTTCATCCGGAAAATGAGACC
ATAAACTCATCGAGGGAAGCAAATGAGTCAAATCTCTCGGATTCTGCAGTTACAAGTATG
GATTACTTTTCTAAGTTCGTCGGCTTATTCTCCTGGTGGCATGGTCATGCCTATCAAGTGG
AATGCAGCAGCAATGGATATTGGCTGCTGCAAACCTTAATATATGATCAGCAGATAGGGGA
CAAGACATGATTGGTCACCAGTCCTTTGTCTTGTCCCTTATCTTTCAGCCAAACGGAAA
GAGAACTTGTGTCTTGGAAAAAAGACATTGAGTTTCTTGGTTTATAAGATTGGTCCTTT
TACCATCCGTTTGGCTGTAAACAGGCAAATCATCTTGGCTCATGCTTCATCAAGTTCTT
ATCTTCGTCGTCTTCTTCTACGCATCTTCATAAGATCTCTGAACTAGTGAATAACATTT
CCTAGCATCATGTTTCAACTAGTGTGTGTGTGAAGAACTCTGCCTTATTTCCAGATGAT
GTATTGTGTGTAACTGTTTATGAAACAAACGTAAGACTTTCAGTTAAAAA

>G1334 Amino Acid Sequence (domain in AA coordinates: 18-190)
MQTELLSPQPWPWNFAFGSQPLTTESLSGEASDSFTGVKAVTTEAEQGVVDKQSTTLF
TFSPGGEKSSRDVPKPHVAFAMQSACFEFGFAQPMMYTKHPHVEQYGVVSAYGSQRSSG
RVMIPKMETEEDGTIYVNSKQYHGIIRRRQSRKAELSRCKPYMHHSRHLHAMRRPR
GSGGRFLNKTADAQSKSPNSQSSEVFHPENETINSSREANESNLSDSAVTSMDYFLS
SSAYSPGGMVMPKWNAAAMDIGCCCKLNI*

>G1381 (32..802)
CAGCTTTAACAATACTCTCTCTCTCTCAAATGGGAAAACAAATCAACATAGAGAGTAG
TGCTACTCATCATCAAGACAATATTGTTTCCGTTATAACAGCCACGATATCCTCCTCCTC
CGTCGTAACGCTTTCGTCAGACTCTTGGTCTACCTCCAAAGATCGTTAGTGCAAGACAA
TGACTCCGGAGGGAACCGCGGAAGAGCAACGTTAGTGATGATAACAAGAATCCGACGTC
GTATAGAGGAGTGAGGATGAGGAGTTGGGAAAATGGGTGTCGGAGATTAGAGAGCCGAG
GAAGAAATCAAGAATATGGCTTGGCACTTATCCAACGGCAGAGATGGCAGCTCGTGCTCA
TGATGTGGCGGCTTTAGCTATTAAAGGCAACTCCGTTTTCTTAATTTCCCTGAATTATC
CGTTTTGCTTCTCGTCCGTTAGCTGCTCTCCTAAGGATATACAAGCTGCAGCTACCAA
AGCCGCCGAAGCAACACGTCGACAAAACCGGTTATCGATAAGAAATTAGCTGATGAGCT
AAGCCACTCTGAGTTGTTGTCTACCGCTCAGTCTTCGACTTCTAGTAGTTTTCGTGTTTC
TTCGGACACGTCGGAGACTTCTAGTACGGACAAGGAAAGCAACGAAGAGACGGTGTGTTGA
TTTGCCGGACCTTTTCACGGACGGGCTTATGAACCCAAACGATGCGTTTTGTTTATGCAA
CGGCACCTTTACGTGGCAGCTTTACGGAGAGGAGGATGTAGGTTTCAGGTTTGAAGAGCC
GTTTAATTGGCAAAATGACTAAACCGCCCTCCACTTGCTTACTGTAATTACTAACATATA
ATTTTCTTGATAAAGACATATATTTCCATTACGGTATTAACATACTTTTCTATCCTTT
TCTCTTTTCTTGTTTCTACATCTGAGTATATTGTCACTATGTGAAAAAATTGATCTCGTT
TTGAATATTTACTTTTCAAATTGAAGTAACGCAAGTGATTGATAAAAAA

>G1381 Amino Acid Sequence (domain in AA coordinates: TBD)
MGKQINIESSATHHQDNIVSVITATISSSSVVTSSSDSWSTSKRSLVQDNDSGGKRRKSN
VSDDNKNPTSYRGVVRMRSWGKWVSEIREPRKKSRIWLGTPTAEMAARAHDAALAIKGN
SGFLNFPPELGLLPVSCSPKDIQAAATKAAEATTWHKPVIDKKLADELHSELLSTAQ
SSTSSSFVFSSTSETSSTDKESNEETVFDLPDLFTDGLMNPDAFCLCNGTFTWQLYGE
EDVGFRFEFPFNWQND*

>G1382 (90..1763)
CTCTCATTTCCGCTAGCTGAGAGCTTCTTCTACTTTCCCTTAGCTTCTTTTTTCTCTCA
TTTTTGTCTTACCCTTGCGAATCTCTGAAATGAACCTCAAGCTAATGACCGGAAGGAGT
TTCAGGGAGATTGTTTCGGCGACGGGAGATCTCACGGCAAAGCACGATTAGCTGGAGGAA
ACGGAGGTGGAGGTGCTAGGTATAAGCTGATGTACCGGCCAAGCTCCGATCTCGAGGT
CGACTGATATCACGATTCCTCCTGGGTTGAGTCCGACTTCGTTTTTGAATCTCCTGTTT
TCATCTCCAACATCAAGCCAGAACCTTCCCTACTACTGGTTCTTTGTCAAGCCTCGAC
CAGTGACATTTCTGCTAGCTCAAGTTCTTATACAGGCAGGGGGTTCCATCAGAACACCT
TTACTGAGCAGAAGTCCAGTGAATTTGAGTTCAGACCTCCTGCATCAAATATGGTATATG
CAGAGCTTGGCAAGATTAGAAGTGAGCCACCACTACATTTTCAAGGCCAGGGCCATGGAT
CCTCACACTCACCTTCTTCAGTCAGTGTGTCAGGTTTCCTCAAGTGAGCTAAGCCGGC
CAACTCCTCCTTGTGATGACACCAACGAGCTCAGATATTCCGGCTGGATCTGATCAAG

AGGAATCAATCCAGACTTCCCAAAATGACTCCAGAGGAAGCACTCCATCCATCTTGGCTG
ATGATGGTTATAACTGGAGAAAATATGGTCAAAAGCATGTCAAAGGGAGTGAATTTCCCC
GGAGCTATTATAAATGTACACATCCTAATTGTGAAGTGAAAAAGTTATTTGAAAGATCTC
ATGATGGGCAGATCACCGATATTATATACAAGGGTACACATGACCATCCTAAACCTCAAC
CTGGTCGCCGAAACTCTGGTGGTATGGCTGCACAAGAAGAAAGGCTAGACAAGTATCCTT
CTTCAACTGGCCGAGATGAGAAGGGATCTGGCGTCTACAACCTTGTCTAACCCCAATGAAC
AAACTGGTAACCTGAAGTACCTCCTATCTCAGCATCTGACGATGGTGGAGAAGCGGCAG
CGTCAAATAGGAATAAAGATGAGCCGGACGATGATGATCCATTCTCAAAACGGAGGAGGA
TGGAGGGTGCATGGAATAACTCCACTAGTGAAACCCATCCGGGAGCCTCGGGTTGTTG
TTCAAACCTCTGAGTGAGGTTGACATTCTGGATGATGGTTATAGATGGCGCAAATATGGGC
AGAAAGTCGTAAGGGGGAACCCAAATCCCAGGAGCTACTACAAATGCACAGCTCATGGAT
GCCCAGTGAGAAAACACGTGGAGAGAGCATCACATGATCCAAAAGCTGTAATAACAACAT
ACGAAGGCAAAACAGATCATGATGTTCCCACTTCAAAGTCTAGCAGCAATCACGAAATCC
AGCCTCGGTTCCAGACCAGATGAAACAGACACCATCAGCCTCAATCTTGGTGTGGAACTCT
CATCTGATGGACCTAACCACGCTTCCAACGAACATCAGCACCAGAATCAACAACCTTGTC
ACCAAACCTACCCAAATGGAGTCAATTTTCAAGTTTGTTCATGCTAGTCCCATGTCATCCT
ACTATGCTAGCTTAAATAGCGGTATGAATCAGTACGGCCAGAGAGAAACAAAGAACGAGA
CTCAAAATGGTGACATCTCGTCCTTGAACAATTCATCTTACCCATATCCGCCCAACATGG
GGAGAGTACAATCGGGTCCGTAAAACAAAAGTAAGCAACATTATGTACGGGATCTTCTT
AGGTTAGGAATGGGACGAGGCTTGTCTATATAATTCCTATTTCTTCACAGAGAGCTGA
TCTTGATTCAAACATATCTCCACATATATTTGTTGTGTACCTGTATTGAGTTCCAA
AAATGTTATGTAAAAATACACAAGATGTTAATGCTTTTATTTAAACAAGAAACAGCA
ATATTACTACAAAAA

>G1382 Amino Acid Sequence (domain in AA coordinates: 210-266, 385-437)

MNPQANDRKEFGDCSATGDLTAKHDSAGNGGGGARYKLMSPAKLPISRSTDITIPPGL
SPTSFLFESPVFISNIKPEPSPTTGSFLKPRPVHISASSSSYTGRGFHQNTFTEQKSSEFE
FRPPASNVMVYAEGLKIRSEPPVHFQGGHSSSHSPSSISDAAGSSSELSRPTPPCQMTPT
SSDIPAGSDQEEISQTSQNSDRGSTPSILADDGYNWRKYGQKHVKGSEFPRSYYKCTHPN
CEVKKLFERSHDGQITDIIYKGTGHDHPKPQPRRNSGGMAAQEERLDKYPSSSTGRDEKGS
GVYNLSNPNQETGNPEVPPISASDDGGEAAAASNRNKDEPDDDDPFSKRMRMEGAMETPL
VKPIREPRVVVQTLSEVDILDDGYRWRKYGQKVVRGNPNPRSYYKCTAHGCPVRKHVERA
SHDPKAVITTYEGKHDHDPVTSKSSSNHEIQPRFRPDETDTISLNLGVGISSDGNHASN
EHQHQNQQLVNQTHPNGVNFVHASPMSSYYASLNSGMNQYQRETNETQNGDISSLN
NSSYPYPNMRVQSGP*

>G1435 (8..904)

GTGAAACATGGGGAAGGAAGTTATGGTGAGCGATTACGGTGACGACGACGGAGAAGACGC
CGGCGCGCGGATGAATATAGGATCCGGAATGGGAAATTGGTTTACCCAACGGAGATGA
TTTGACTCCGTTATCTCAATATCTAGTCCCGTCGATTCTCGCGTTAGCTTTTACGATGAT
CCCAGAACGAAGCCGTACAATTCACGACGTCAATCGCGCGTCGAAATCACGCTCTCTTC
GTTGAGAAGCAGTACCAATGCTTCTGTGATGGAGGAGGTCGTGGATCGAGTTGAATC
GAGTGTTCAGGATCAGATCCGAAGAAACAGAAGAAATCGGATGGTGGTGAAGCAGCGGC
GGTGGAGGATTCCACGGCGGAGGAAGGAGACTCCGGGCCCTGAAGACGCGCTCTGGGAAGAC
ATCGAAACGACCGCGTTTAGTGTGGACACCGCAGCTACACAAGAGATTTGTGGACGTTGT
GGCTCATCTAGGGATTAAAAACGCAGTGCCGAAGACGATTATGCAGCTGATGAACGTGGA
AGGACTTACTCGTGAGAACGTTGCGTCTCATTTGCAGAAATATAGGCTTTACCTTAAACG
GATTCAGGATTGACGACGGAAGAAGATCCTTATTCGTCGTCGGATCAGCTCTTCTCTTC
AACGCCGGTTCTCTCCACAGAGCTTTCAAGACGGCGGAGGAAGTAACGGAAAGTTGGGGGT
TCCGGTTCCGGTTCGTCGATGGTGCCTATTCCAGGCTATGGGAATCAAATGGGTATGCA
AGGATATTATCAACAGTATAGTAACCATGGCAATGAATCAAACCAATATATGATGCAGCA
GAATAAGTTTGGAAACATGGTGACATATCCTTCTGTTGGTGGTGGTGACGTGAATGACAA
GTAAATGGATCTTAAAGGTCTATAATTTGCTCTACAGAGAGATACTGGTCTTGGCTTAT
GGTTTATTTTCCCACTTCATGAGGTTGTTGTGACTTTTAAATTCTCCATGTTTCCACACA
AGTCTTTATTTGCTTTGTATAGAAAATGATTTTCGAGAAAATCACTGGGAAGCTTGGTATT
GTTGGAGGATGAAGCCTTCTATGAATGATTTAGTTTCTTACTGTCTCCATTCTTTATGAG
GTAATAAAGCCTTCTTTGCTCATCGCTTGTAGTCTTCTTAAATTCAAGACAGCGTCACA
TGTTTGTTCGGTTATGTTAATTGTTTCTTTTGGATAATGAAGATAGCATCAGGTCTC
ATGTCTCCTCACTTTGATAAA

>G1435 Amino Acid Sequence (domain in AA coordinates: 146-194)
MGKEVMVSDYGGDDGEDAGGGDEYRIPEWEIGLPLNGDDLTPLSQYLVP SILALAFSMIPE
RSRTIHDVNRASQITLSSLRSSSTNASSVMEVVDRVSVPGSDPKKQKSDGGEAAAVE
DSTAEEDSGPEDASGKTSKRPLVWTPQLHKRFVDVVAHLGIKNAVPKTIMQLMNVEGL
TRENVAHLQKYRLYLKRIQGLTTEEDPYSSSDQLFSSTPVPQSFQDGGGSGNKLGVVPV
PVPSMVPIPGYGNMQGYQQYSNHGNE SNQYMMQONKFGTMVITYPSVGGGDVNDK*
>G1537 (1..783)

ATGGAACCGAAGTAAACGCAGGAACAGCAAGCAGTTCAAGATGGAACCCAAACGAAAGAT
CAGATCACGCTACTGGAAAATCTTTACAAGGAAGGAATACGAACTCCGAGCGCCGATCAG
ATTCAGCAGATCACCGGTAGGCTTCGTGCGTACGGCCATATCGAAGGTAAAAACGTCCTT
TACTGGTTCCAGAACATAAGGCTAGGCAACGCCAAAAGCAGAAAACAGGAGCGCATGGCT
TACTTCAATCGCCTCCTCCACAAAACCTCCCGTTTCTTCTACCCCCCTCCTTGCTCAAAC
GTGGGTTGTGTCACTCCGTACTATTTACAGCAAGCAAGTGATCATCATATGAATCAACAT
GGAAGTGTATACACAAACGATCTTCTTACAGAAACAATGTGATGATTCCAAGTGGTGGC
TACGAGAAACGGACAGTCACACAACATCAGAAACAACCTTTCAGACATAAGAACAACAGCA
GCCACAAGAATGCCAATTTCTCCGAGTTCACTCAGATTTGACAGATTTGCCCTCCGTGAT
AACTGTTATGCCGGTGAGGACATTAACGTCAATTCCAGTGGACGGAAAAACTCCCTCTT
TTTCTCTTTCAGCCTTTGAATGCAAGTAATGCTGATGGTATGGGAAGTTCAGTTTTGCC
CTTGATAGTGATTCTCCGGTGGATTGTTCTAGCGATGGAGCCGGCCGAGAGCAGCCGTTT
ATTGATTTCTTTCTGTTGTTCTACTTCTACTCGTTTTCGATAGTAATGGTAATGGGTTG
TAA

>G1537 Amino Acid Sequence (domain in AA coordinates: 14-74)
MENEVNAGTASSSRWNPTKDQITLLENLYKEGIRTPSADQIQQITGRLRAYGHIEGKNVF
YWFQNHKARQRQKQKQERMAYFNRLLLHKTSRFFYPSPCSNVGCVSPYYLQQASDHMNQNH
GSVYTNDDLHRNNVMIPSGGYEKRTVTQHQQQLSDIRTTAATRMPIPSSSLRFDRFALRD
NCYAGEDINVNSSGRKTLPLFPLQPLNASNADGMGSSSFALGSDSPVDCSSDGAGREQPF
IDFFSGGSTSTRFDSNGNGL*

>G1545 (67..729)
CATCACCAATCTTTTGAATCTAAGAGAGAGAAGAAGAAGGTCTAGAGAACGAAAAGA
AGAAACATGAATAACCAGAATGTAGATGATCATAATCTTCTACTCATTTCTCAATTGTAC
CCTAATGTCTATACTCCATTAGTACCACAACAAGGAGGAGAAGCAAAACCAACACGGCGG
AGGAAAAGGAAGAGCAAGAGTGTGTGGTGGCAGAGGAGGTGAAAACGAAGGCAATGGG
TGGTTTAGAAAGAGAAAATTTGAGTGATGAGCAAGTAAGAATGTTGGAGATTAGCTTTGAA
GACGATCATAAGCTTGAATCCGAGAGGAAAGATCGGCTTGCTTCTGAGTTAGGGCTTGAT
CCTCGTCAAGTCGCCGTCTGGTTCCAAAACCGCCGTGCACGGTGGAAAGAACAAACGAGTC
GAGGATGAATACACTAACTCAAGAATGCATACGAAACCACCGTCGTTGAGAAATGTCTGT
CTTGATTCTGAGGTTATTACCTAAAGGAACAACCTTACGAGGCTGAAAGAGAGATCCAA
CGGCTTGCAAAAAGAGTTGAAGGAACCTTTAAGTAACAGTCCTATCTCATCTCTGTGACC
ATTGAAGCCAATCATACGACACCGTTTTTTGGAGATTACGACATCGGATTTGACGGTGAG
GCTGACGAGAACTTGCTCTACTCGCCAGATTACATTGATGGATTAGACTGGATGAGCCAA
TTTATGTAAAAAACTATAAGCTAATCTATTTTCAGTCGTAGTATAG

>G1545 Amino Acid Sequence (domain in AA coordinates: 54-117)
MNNQNVDDHNNLLISQLYPNVYTPLVPQQGGEAKPTRRRKRKSKSVVVAEEGENEGNGWF
RKRKLSDEQVRMLEISFEDDHKLESEKDRLASLGLDPRQVAVWFQNRARWKNKRVED
EYTKLKNAYETTVVEKCRLDSEVIHLKEQLYEAREIQRLAKRVEGTLSNSPISSSVTIE
ANHPTTFFGDYDIGFDGEADENLLYSPDYIDGLDWSQFM*

>G1641 (1..867)
ATGGAGGTTATGAGACCGTCGACGTCACACGTGTCAGGTGGGAACCTGGCTCATGGAGGAA
ACTAAGAGCGCGTCGCGTCTCTGGTGAAGGTGCCACGTGGACGGCGGCAGAGAACAAG
GCATTGAGAAATGCTTTGGCGGTTTACGACGACAACACTCCTGATCGGTGGCAGAAGGTG
GCTGCGGTGATTCCGGGGAAGACAGTGAGTGACGTAATTAGACAGTATAACGATTTGGAA
GCTGATGTCAGCAGCATCGAGGCCGTTTAAATCCCGGTCCCGGTTACATCACCTCGCCG
CCTTTCACTCTAGATTGGGCCGGCGCGGTGGCGGATGTAACGGGTTTAAACCGGTCAT
CAGGTTTGTAAATAACGGTCGCAGGCCGGTAGATCGCCGAGCTGGAGCGGAAGAAAGGC
GTTCTTTGGACGGAGGAAGAACAAGCTATTTCTAATGGGTTTGAAGAAATATGGGAAA
GGAGATTGGAGAAAACATATCTCGGAACCTTTGTGATAACGCGAACGCCAACACAAGTAGCT
AGCCACGCCCCAAAAGTACTTCATCCGCAACTTTCCGGCGCAAGGACAAGAGACGAGCA

AGCATTACACGACATAACCACCGTAAATCTCGAAGAGGAGGCTTCTTTGGAGACCAATAAG
AGCTCCATTGTTGTTGGAGATCAGCGTTCAAGGCTAACCAGCGTTTCTTGGAAACCAAACG
GACAACAATGGAACACAGGCAGACGCTTCAATATAACGATTGGAAACGCTATTAGTGGC
GTTTCATTACATACGCCAGGTTATGATTGGAGGGTATAACAATGCAGATTCTTGCTATGAC
GCCCAAAACACAATGTTTCAACTATAG

>G1641 Amino Acid Sequence (domain in AA coordinates: 139-200)
MEVMRPSTSHVSGGNWLMEETKSGVAASGEGATWTAENKAFENALAVYDDNTPDRWQKV
AAVIPGKTVSDVIRQYNDLEADVSSIEAGLIPVPGYITSPPTLDWAGGGGGCNGFKPGH
QVCNKRSQAGRSPELERKKGVPTWEEHKLFLMGLKKGKGDWRNISRNFVITRPTQVA
SHAQKYFIRQLSGGKDKRRASIHDIITVNLEEEASLETNKSSIIVVDQRSRLTAFPNQOT
DNNGTQADAFNITIGNAISGVHSYQVMIGGYNNADSCYDAQNTMFQL*

>G165 (19..699)
CTTCAAAACATCTAAAAAATGGTGAAAAAACTCTTGGTCGTAGAAAGGTAGAGATAGTG
AAAATGACTAAGGAATCAAACCTTCAAGTCACATTTTCCAAGAGAAAAGCTGGTCTTTT
AAGAAGGCTAGTGAATTTTGCACATTATGTGATGCAAAAATTGCGATGATCGTGTTC
CCAGCTGGAAAAGTATTTTCTTTTGGTCATCCAAATGTTGATGTTCTGCTTGACCACTT
CGAGGGTGTGTTGTAGGACACAACAACACAAACCTTGATGAAAGCTACACAAAGCTTCAT
GTTCAAATGCTCAACAAATCCTACACTGAGGTGAAGGCGGAAGTAGAAAAAGAACAAAAG
AATAAGCAGTCGCGGGCTCAAATGAAAGAGAAAACGAAAACGCTGAGGAGTGGTGGAGT
AAGTCTCCATTAGAACTCAACTTAAGTCAATCAACCTGTATGATACGTGTTCTTAAAGAT
TTGAAGAAGATAGTTGATGAAAAAGCAATTCAATTAATCCATCAACAAACCCAACTTC
TATGTTGGAAGTTCTAGCAATGCTGCTGCTCCAGCAACTGTTAGTGGTGGTAATATCTCC
ACAAACCAGGGGTTCTTGTATCAAAACGGAATGACGACTAATCCTACTCAAAACACTTCTG
TTTGGATTTGATATTATGAATCGCACACCAGGAGTTTAAATAAGTCTATCCTCATTATGG
GTCTTGGTACTATAAGTTCATCTCTCTCGTTGTTGACTTTTTTAAGTCTCCAATAGTTTGT
TGTG

>G165 Amino Acid Sequence (conserved domain in AA coordinates:7-62)
MVKKTLGRRKVEIVKMTKESNLQVTFSSKRKAGLFKKASEFCTLCDAKIAMIVFSPAGKVF
SFGHPNVDVLLDHFRGCVVGHNNNTNLDES YTKLHVQMLNKS YTEVKA EKEQKNKQSR
QNERENENAE EWSKSPLELNL SQSTCMIRVLKDLKKIVDEKAIQLIHQTNPNFYVGS
SS NAAAPATVSGGNISTNQGFDFDQNGMTTNPQTLLFGFDIMNRTPGV*

>G1652 (77..1078)
AGCAAGTCCAAATCTCCCTCTCTCTCTCTATCTATCTCTCTATAGAAGATTTTTTAAC
TAAGAAGCTAGCGATCATGGCCACAGCGATGAACGTTTTCTCTACCAAATGGTCTCCGA
ATTGGATATAGAAGAATATAGTATCATCCACCAATTCACATGAACCTCACTCGTCGGAGA
TGTTCCACAGTCTCTCTCATCTCTTGATGATACCACCACTTGTTATAACCTTGATGCTTC
TTGTAATAAAAGTTTGGTGAAGAAAGACCTTCAAAGATCCTCAAGACCACTCACATATC
ACCAAACCTTACATCCTTTTTCTTCTTCTAATCCCTCCTCCTCAAAGCACCAGCCCTCTTC
TAGGATCTTTCTTTTGAAGACAGGTTTACATGTTATGAATCACAACCTCTCCAACTT
AATATTTAGCCCCAAGGACGAAGAAATTGGATTACCAGAGCATAAGAAAGCCGAGCTGAT
AATAAGAGGGACAAAGAGAGCTCAATCCTTGACTCGAAGCCAATCAAATGCTCAAGATCA
CATACTGGCAGAGAGAAAACGGAGAGAGAAGCTTACTCAAAGATTTGTAGCTCTTTCCGC
GCTAATCTCTGGCCTAAAGAAGATGGACAAGGCTTCTGTGTTGGGAGATGCAATAAAGCA
TATAAAGTACCTCCAAGAGAGTGTGAAAGAGTATGAGGAACAAAAGAAGGAAAAGACAAT
GGAATCAGTGGTTCTTGTAAAGAAGTCTAGTCTGGTTTATAGATGAAAATCATCAACCATC
ATCATCATCTTCTCAGATGGAATCGCAATAGCTCGAGCTCAAATCTTCCAGAAATAGA
AGTTAGGGTTTCAGGAAAAGATGTTCTTATTAAGATCCTATGCGAGAAGCAAAAGGGTAA
TGTGATCAAGATTAAGGGGGAGATTGAAAAGCTTGGTTTGTCTATCACCACAGCAATGT
CTTGCCCTTTGAGCCCACTTTTGACATCTCTATTATCGCTCAGAAGAATAACAATTTTGA
TATGAAAATCGAGGATGTTGTGAAGAACTTGAGTTTGGCTTATCAAAGCTCACTTAATT
GGTTTTCAGTTACATACATATACACATTATCATCGATTCTCCGATCGAAGAATCCAAA
ATCAGTTTTTCCATGAAAGTGGTTTTTGTAGTTGTTAAGTTTGGTTGTATGGAGATTCTTAA
GTCATTTAAAGATCCTTGTCTTGTGTTGTTAAGTGTGCTTTAAGATGCATATCATCAA
TGTTTTAGTAATTATTTCTCTCCAGTTTCATTTGGGACGGAATTTTTTTCGCAGTTGTTGG
ATATATATTTCTCGCATGTAAAGCATTTCTGTTAGTTTAAATAAACGTCCGATATGTTTCT
TTGAAAA

>G1652 Amino Acid Sequence (domain in AA coordinates:143-215)

MATAMNVFSTKWSELIDIEEYSIIHQFHMNSLVGDVPQSLSSLDLDDTTTCYNLDASCNKSL
VEERPSKILKTTHISPNLHPFSSSNPPPKHQPSRILSFEKTGLHVMNHNSPNLIFSPK
DEEIGLPEHKKAELIIRGTRKRAQSLTRSQSNADHILAERKRREKLTQRFVALSALIPGL
KKMDKASVLGDAIKHIKYLQESVKEYEYEQKKEKTMESVVLVKSSSLVDENHQPSSSSSS
DGNRNSSSSNLPEIEVRVSGKDVLIKILCEKQKGNVIKIMGEIEKLGLSITNSNVLFPFGP
TFDISIIAQKNMNFDMKIEDVVKNSLSPGLSKLT*

>G1655 (132..755)

TTTCTAACTAGTCACATTGAGAGAGAGAGAGAGAGAGAAAGAGAGACTCTCAGAATCTGAAG
AAGAAGAAGAGATTGTTGTTTTTGCCTTTTATCATCGGTTTCTTTGAATCTCTGGTTTTA
AATCGGATTTAATGGTGGAGTCTCTGTTCCCGAGCATCGAAAACACAGGTGAATCGTCTC
GAAGAAAGAGCCGAGGATATCAGAGACGGCGGAGGCGAGATAGAGGCACGACGTGTCA
ACGAAGAAAGCTTGAAGAGATGGAAGAACGAATCGTGTGCAACAGATCTACGTTGTAAAGC
TCGTGCAAGCTTTACGCCGAGTTCGTGAGAGATCTTCCACCACCAGCAACACGAGACCG
ATAAATCGTCTCCGGCGCGGCGAGGGAGATACGTGATACGGCGGATCGAGTTCTAGCTG
CGTCCGCTCGTGGTACGACTCGGTGGAGCAGAGCGATTTTAGCGAGTCGCGTCCGAGCGA
AGCTGAAGAAACATAGAAAGGCGAAAAAGTCAACGGGAAATTGTAAATCGAGAAAAGGTC
TCACGGAGACGAATCGGATTAAGTTACCGGCGGTTGAGAGAAAAGTGAAGATTCTTGGCC
GTTTGGTTCTGTTGTTGCCGAAAGTCTCTGTACCGAATCTTTTAGATGAAGCGACCGATT
ACATCGCAGCGTTAGAGATGCAGGTTGAGCCATGGAGGCTCTCGCCGAATTTTAACCG
CAGCCGACCCACGGACGACGTTGACCCGAACTTAACGGCGGAGTTAGTTTGTAGTTGT
TAATTAGCTTTTCTTTTACCTTTTACCCCTTTATTTTGGCTTCAAGTGTTTTTTTTTTTC
TCGTCGACGCGATTTTAATTTATTAATTC

>G1655 Amino Acid Sequence (domain in AA coordinates: 134-192)

MVESLFPSIENTGESSRRKKPRISETABAEIEARRVNEESLKRWKTNRVQIYACKLVEA
LRRVRQRSSSTSNNETDKLVSGAAREIRDTADRVLAASARGTTRWSRAILASRVRAKLLK
HRKAKKSTGNCKSRKGLTETNRIKLPAVERKLKILGRLVPGCRKVSVPNLLDEATDYIAA
LEMQVRAMEALAEELLTAAAPRTTLTGT*

>G1671 (188..751)

TCCCACTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTTCATTGGAGAGG
ACACGCTGACAAGCTGACTCTAGCAGATCTGGTACCGTCGACCCTCTCTATATAATCTTC
TTCTACACACACACACACGCAACCATATACGTACATGTGAAGTAGTGAGATCAATATC
GTTAGCAATGAATCTACCACCGGATTTAGGTTTTTTCCGACCGATGAAGAGCTCGTCGT
TCACTTCCTCCACCGGAAAGCTTCCCTCTTGCCTTGTCAACCTGATGTCAATCCCCGACCT
TGATCTTTACCATTACGATCCTTGGGACCTTCCCGGAAAGCTTTGGGAGAAGGGAGGCA
ATGGTACTTCTATAGTAGAAAGACACAAGAGAGAGTGACAAGCAATGGGTATTGGGGATC
AATGGGAATGGACGAGCCAATCTACACAAGCTCCACACACAAGAAAGTGGGAATCAAAAA
GTATCTAACTTTCTATCTCGGAGATTCTCAGACTAATTGGATCATGCAAGAAATATTCCCT
CCCGGATTCTCTTCTTCTATCTAGTCGATCTTCTAAGAGATCAAGCCGTCTCTAGTTC
TAGTCACAAACCCGATTATAGCAAGTGGGTGATATGCAGAGTGTATGAGCAAAATTGCAG
TGAGGAGGAAGAGGATGATGGGACAGAAGTCTCATGTTTGGATGAAGTGTTTTTGTCTTT
AGATGATCTTGACGAAGTAAGCTTACCGTAATAAAGACAGAAGCACCCAAGAAGAGAAAA
AAAAAAAAGGGTTTAGTGGGCAATTATTTCTAAGCGACCGCTCTAGACAGGCCTAGTAC
CGGATCTCTAGCTAGAGCTTTCGTTTCGTATCATCGGTTTTCGACAACGTTTCGTCAAGT

>G1671 Amino Acid Sequence (domain in AA coordinates: TBD)

MNLPPGFRFFPTDEELVVHFLHRKASLLPCHPDVIPDLDLHYHDPWDLPGKALGEGRQWY
FYSRKTQERVTSNGYWGSMGMDPIYTSSTHKKVGIKKYLTFFYLGDSQTNWIMQEYSLPD
SSSSSSSRSSKRSSRSSSHKPDYSKWVICRVYEQNCSEEDDDGTELSCLDEVFLSLDD
LDEVSLP*

>G1756 (71..1003)

ATATGTACTTGTACACCAACCCACCAAAAGAGATAAAAGAGGAAACAAAAACTCGAAAAG
AGAGAGATATATGGGTGAGGTGGCTTATATGGACGAAGGAGACCTAGAAGCAATAGTCAG
AGGCTACTCCGGCTCCGGAGACGCGTTTTCCGGCGAAAGTTCGGTACGTTTTTACCTTC
GTTTTGCCTACCGATGGAGACGTCTAGTTTCTACGAACCGGAGATGGAGACAAGTGGCTT
AGATGAGCTCGGTGAACTTTACAAACCTTTTACCTTTTCTCCACACAACGATCCTCAC
AAGCTCGGTCTCTCTCCCTGAAGATTCAAAACCTTTCCGAGATGACAAGAAAACAACGATC
ACATGGTTGTCTTTTATCCAACGGATCAAGAGCTGATCATATCCGAATTTCAGAATCCAA
ATCAAGAAAAGCAAGAAGAAATCAACAGAAGAGAGTGTGTTGAGCAAGTGAAAGAAGAGAA

TCTGTTGTGCGACGCATGGGCGTGGCGTAAATACGGGCAGAAACCCATCAAAGGATCTCC
 ATACCCAAGGAGTTATACAGATGCAGTAGCTCAAAAGGGTGTGTTGGCAAGAAAACAAGT
 CGAAAGAAATCCTCAAAACCCGGAGAAATTCACCATAACATACTAATGAGCACAAATCA
 TGAATACCAACCCGGAGAAACTCATTAGCCGGTTCGACTCGAGCAAAAACCTCCCAACC
 CAAACCAACCTTAACCAAAAAATCCGAAAAAGAGTTGTTTCTCCCTACAAGTAATCC
 TATGATCCCATCCGCTGATGAATCTTCTGTTGCGGTTCAAGAAATGAGCGTTGCGGAAAC
 GAGTACGCACCAAGCGGCTGGAGCAATCGAGGGCCGCGCTTGAGTAACGGTTTACCATC
 GGATTTGATGTCCGGGAGCGGAACTTTTCCAGTTTTACCAGTGACTTCGATGAATATT
 GAATAGCCAAGAGTTCTTCAGTGGGTATTTATGAATTACTAGAGAGCATTAGGTGTATG
 TATATATATAT

>G1756 Amino Acid Sequence (domain in AA coordinates: TBD)
 MGEVAYMDEGDLEAIVRGYSGSDAFSGESSGTFSPSFCLPMETSSFYEPMETSGLDEL
 GELYKFPYFSTQTILTSSVSLPEDSKPFRDDKKQRSHGCLLSNGSRADHIRISESKSKK
 SKKNQQRVVEQVKEENLLSDAWARKYQKPIKGSYPYRYSYRCSSSKGLARKQVERN
 PQNPEKFTITYTNEHNHELPTRRNSLAGSTRAKTSQPKPTLTKKSEKEVVSSPTSNNMIP
 SADESSVAVQEMSVAETSTHQAGAIEGRRLSNGLPDLMSGSGTFPSFTGDFDELLNSQ
 EFFSGYLWNY*

>G1757 (250..1224)
 ATACCAATCCTATAACACTCTCATTTCTCATCATATCATTCTTCAATCTATATAACCCAT
 TCTTAATTATACTCAACACACATTATATTTTTCTGATCATATCATTCTTTCAGTCCATCT
 ATATAACCAATTCTTGATTTATACTTAAACACACATTATACATCTTTCTCATCATAGTT
 TGTATCAATTTCTAGAGTAACTACCTAAAGGAAAAAAAATCTATTTTGGGAATCAT
 ATACTAAAAATGGAAGGAAGAGATATGTTAAGTTGGGAGCAAAAGACATTGCTAAGCGAG
 CTTATCAATGGATTTGATGCGGCCAAAAAGCTTCAGGCACGACTTAGAGAAGCTCCGTGCG
 CCGTTCGTCACTCAATTTTATCACCAGCGACGGCTGTTGCTGAGACTAACGAGATTCTGGTG
 AAGCAGATAGTTTCTTCTACGAGAGATCTCTTCTTCTGCTAAACTGGTCATCCTCACC
 AGCGTACAACCTATTCCGACGCCGTTACTGTAGTCCCGGTGGCAAATCCCGGCAGTGTT
 CCAGAATCTCCGCATCGATAAACGGAAGTCCGAGAAGTGAAGAGTTTGCCGATGGAGGA
 GGTTCTAGCGAGAGTCATCATCGCCAAGATTACATTTTCAATTCAAAGAAAAGAAAGATG
 TTACCAAAGTGGTCAGAAAAAGTGAGAATAAGCCCAGAGAGAGGCTTAGAAGGACCTCAA
 GATGATGTCTTTAGCTGGAGAAAATATGGTCAAAAAGACATTTTAGGCGCCAAATTCCCA
 AGGAGTTATTACAGATGCACACATCGTAGCACACAAAACCTGTTGGGCAACGAAACAAGTC
 CAGAGATCAGACGGGGATGCTACGGTTTTTCGAAGTGACGTACAGAGGAACACACACTTGT
 TCGCAGGCGATCACAAGAACACCACCATTAGCCTCGCCGAGAAGCGACAAGACACCAGA
 GTCAAACCAGCCATTACCCAAAAGCCAAAGGATATTCTCGAGAGTCTTAAATCCAACCTTA
 ACCGTTTGAACCGATGGGCTTGATGATGGTAAAGACGTTTCTCGTTCCCTGTACGCCG
 CCGTTTTACAATTACGGAATATCAACGGCGAGTTCGGCCACGTGGAGAGTTCTCCGATC
 TTCGACGTTGTTGACTGGTTCAATCCAACGGTCGAGATTGACACAACCTTCCCCGCGTTT
 TTACACGAGTCGATTTATTATTAATTAATAATTTGTAACAGAGAAATAGATAGTAAGT
 AAGTAATGATCAGCGAGAGTTAAACATAAAAAGTACTTAGAGTAATCTAACGATGCATAA
 TAAGGAATGTTTCAACAGGACTTGAACATGATTTCAATACTAAGAGAGATTTATCTAGCTA
 CTGGTAGTAGCCGCAGACTTCTTGTGTAGCTTCACTTNCCTTTTGTGCTT

>G1757 Amino Acid Sequence (domain in AA coordinates: 158-218)
 MEGRDMLSWEQKTLLELSELINGFDAKKLQARLREAPSPSSSFSSPATAVAETNEILVKQI
 VSSYERSLLLLNWSSSPSVQLIPTVTVVPVANPGSVPEPASINGSRSEEFADGGSS
 ESHHRQDYIFNSKKRKMPLKWSEKVRISPERGLEGPQDDVFSWRKYQKQDILGAKFPRSY
 YRCTRSTQNCWATKQVQRSDGATVFEVTRYRGTHTCSQAITRTPPLASPEKRQDTRVKP
 AITQPKPDILESLSNLTVRTDGLDDGKDVFSFPDTPPFYNYGTINGEFHVESSPIFDV
 VDWFNPTVEIDTTFPAFLHESIYY*

>G1782 (1..927)
 ATGCAAGTGTTCAAAGGAAAGAGATTATCTTGGGGAACTCAATGCCTACAACAAAT
 TCAAATATTCAAGGATCTGAATCTTTCAGCTTGACTAAGGATATGATAATGTCTACAACA
 CAATTACCCGCGATGAAACATTCGGGTTTGACGCTGCAAAATCAAGATTCAACCTCATCA
 CAATCTACTGAAGAAGAATCAGGCGGCGGTGAAGTTGCAAGCTTTGGAGAATATAAGCGT
 TATGGATGCAGCATTGTTAATAACAATCTCTCAGGTTACATCGAAAACCTTGGGAAAGCCT
 ATTGAAAATTATACTAAGTCAATTACTACCTCGTCGATGGTGTCTCAAGACTCTGTGTTT
 CCTGCTCTACTTCTGGTCAAAATATCTTGGTCTCTTCAATGTGCTGAAACGTCAATTTT

AATGGTTTCTTGGCTCCTGAATATGCATCAACACCAACGGCGCTGCCACATTTAGAGATG
ATGGGTTTGGTTTCTTCAAGAGTGCCATTGCCTCATCACATTCAAGAGAATGAACCAATA
TTTGTCAATGCGAAACAGTATCATGCGATTCTCCGTCGCAGGAAGCACCCTGCTAAACTC
GAAGCTCAGAACAAACTCATCAAATGCCGTAAACCGTACCTTCATGAGTCTCGCCATCTT
CATGCTTTAAAGAGAGCTAGAGGCTCCGGTGGACGTTTCTCAATACAAAGAAGCTTCAA
GAATCATCAAACCTACTGTGTTCTTCTCAAATGGCAAATGGACAAAATTTCTCTATGAGC
CCTCACGGTGGTGGAGCGGAATCGGGTCTAGTTTCGATCTCACCGAGCTCCAATTCAAAC
TGTATCAACATGTTCCAAAACCCGCAGTTCAGATTCTCAGGTTATCCGTCAACACACCAT
GCCTCAGCTCTCATGTGAGGGACTTGA

>G1782 Amino Acid Sequence (domain in AA coordinates: 166-238)

MQVFQRKEDSSWGNMPTTNSNIQSESFSLTKDMIMSTTQLPAMKHSGLQLQNQDSTSS
QSTEEESGGGEVASFGEYKRYGCSI VNNLSGYIENLGKPIENYTKSITTSSMVSQDSVF
PAPTSGQISWSLQCAETSHFNGFLAPEYASTPTALPHLEMMGLVSSRVPLPHHIQENEP
FVNAKQYHAILRRRKHRAKLEAQNKLKCRKPYLHESRHLHALKRARGSGGRFLNKKLQ
ESSNSLCSSQMANGQNFMSMPHGGSGIGSSSISPSSNSNCINMFQNPQFRFSGYPSTHH
ASALMSGT*

>G184 (327..1937)

TGAATTCTAGCCTTTTGTAGGCGAATCATCTGGACCGGTAAGAGACTCTCTCATCGATA
ATAACCACATAATTTAATCAAACCTCTTCTCTCTCTTCTAAGATCTTTTGCTTTGCTCT
TTTCTCTTTTGGATCTTCTATATATGGAGAAGCACCAAAACGGTACTTACTATACGATAC
TGTACGGATCCATCAAACCTGGATTAATTATCAAAACGTACATTTTATCTTACCTGGCAA
GTTACATTCTTAGGGTTTGGAGAATCCAATCAACAACAAAGAAAATAATCATCGTTACA
ATAATCAGTATCACGCACAGACTTAGATGTTCCGGTTTCCAGTGAGTCTAGGCGGTTTAC
GTGACGAAGACCGTCACGATCAGATCACACCGTTGGATGACCATCGTGTGGTGGTTGATG
AGGTTGACTTCTTCTCAGAGAAGAGAGATAGGGTTTACGTTGAGAATCAACGACGACG
ACGACGAAGGCAATAAGGTTCTCATCAAATGGAGGGTTACGAGTTGAAGAAAACGATC
GTTCCAGAGATGTCAATATCGGTCTGAATCTTCTGACCGGAATACGGGAAGCGATGAGT
CAACGGTGGATGATGGACTATCAATGGATATGGAAGATAAACGTGCAAAGATTGAGAACG
CACAACCTACAAGAAGAGCTCAAGAAGATGAAAATAGAGAATCAAAGGCTAAGAGATATGT
TGAGCCAAGCGACGACCAACTTCAATGCCTTACAAATGCAACTTGTTGCCGTCATGAGGC
AACAGAACAACGTAACCTTTCACAAGATCATCTCCTGGAGAGCAAAGCAGAAGGAAGGA
AACGGCAGGAACCTGCAATCATGGTGCCAAGGCAGTTCATGGACCTTGGGCCGTCTGTCTG
GAGCAGCAGAGATGGAGCCGAAGTGTCTATCTGAAGAGAGGACAACGGTTCTGTTCAAGTT
CTCCTCCTTCGCTTCTAGAAAGTTTCCAATCCCCGAGAGAACGGAAAGAGGTTGCTTGGAA
GAGAAGAAAGCTCAGAGGAATCAGAGTCTAACGCCTGGGGAAACCTAACAAAGTCCCCA
AACATAATCCATCCTCTAGCAATAGCAATGGAACAGAAACGGAAATGTTATGATCAGT
CGGCCGAGAAAGCCACCATGCGGAAAGCCCGTGTCTCAGTTCTGTGCCCGATCTGAAGCTG
CCATGATAAGCGATGGATGTCAATGGAGAAAGTACGGACAAAAAATGGCTAAAGGAAACC
CGTGTCCGCGGGCTTATTATCGTTGCACAATGGCCGGTGGATGTCCAGTTTCGCAAGCAAG
TGCAGCGTTGCGCAGAAGACAGATCTATTCTCATAACCACCTACGAAGGAAACCACAACC
ATCCACTCCCACAGCCGCTACGGCCATGGCCTCAACAACCACCGCAGCTGCAAGCATGC
TCCTCTCGGGCTCAATGTCTGAGTCAAGACGGTTTAAATGAACCCAACAAACCTCCTAGCTC
GAGCTATCTTGCCTTGCTCTCAAGCATGGCTACAATCTCAGCCTCCGCACCATTTCCCAA
CCATCACATTGGACCTCACCAATTCACCAACGGTAACAACCTAATATGACCACTAATA
ACCCGTTGATGCAGTTTCGCTCAACGGCCCGGTTTCAACCCGGCAGTTTTCGCTCAAGTGG
TTGGTCAAGCTATGTACAATAACCAACAACAGTCCAAGTTTCTGGTTTACAGTTACCGG
CTCAGCCACTGCAGATCGCGGCCACTTCCTCGGTGGCCGAGAGCGTTAGTGTGCCAGTG
CAGCAATTGCGTCCGATCCAACTTTGCGGCGGCTCTAGCGGCAGCGATCACGTCCATTA
TGAACGGTTCCAGTCATCAAATAATAACACCAATAATAATAATGTGGCTACGAGCAACA
ATGACAGTAGGCAATAAGAGTTTTTCAATTTGATGGTCGATTTTTTTTTTTGGGG

>G184 Amino Acid Sequence (domain in AA coordinates: 295-352)

MFRFPVSLGGSRDEDRDQITPLDDHRVVVDEVDFSEKRDRVSRENINDDDDDEGNKVL
KMEGSRVEENDRSRDVNI GLNLLTANTGSDESTVDDGLSMDMEDKRAKIENAQLQEELKK
MKIENQRLRDMLSQATTNFNALQMLVAVMRQQEQRNSSQDHLLESKAEBGRKRQELQIMV
PRQFMDLGPSSGAAEHAESSEERTTVRSGPSPLLESSNPRENKRLLGREESSESE
SNAWGNPNKVPKHNPSNSSNNGNRNGNVIDQSAAEATMRKARVSVRARSEAAMISDGCQW
RKYGQKMAKGNPCPRAYRCTMAGGCPVRKQVQRCEDRSILITTYEGNNHNLPLPPAATA

MASTTTAAASMLLSGSMSSQDGLMNPTNLLARAILPCSSSMATISASAPFPITLTLTNS
 PNGNNPNMTTNNPLMQFAQRPGFNPVLPQVVGQAMYNQSQSKFSGQLQPAQPLQIAAT
 SSVAESVSAASAAIASDPNFAAALAAAITSIMNGSSHQNNTNNNNVATSNNDNRQ*

>G1845 (111..989)

AAGACATAATTTTCTCTGTTTTCTAGCTCTCTCCTCTCAAATTTCTTCCATTGCTCTCTG
 TTTTGGCAAATCGTGAAGTCCACGCTCTTTAAGGCATCAGTGAAGCAAAGATGGACTTTG
 ACGAGGAGCTAAATCTTTGTATTACGAAAGGTAAAAATGTTGATCATTCTTTTGGAGGAG
 AAGCTTCTTCCACGTCCCCAAGATCTATGAAGAAAATGAAGAGTCTAGTCGTCCTAAAC
 CCTATTTCCAATCCTCTTCTCTCTTATTCGTTAGAGGCTTCCCTTTTTCTCTCGATC
 CAACACTTCAGAATCAGCAACAACAACTCGGATCATACTTCCGGTACTTGAGCAACGAC
 AAGACCCGACAATGCAAGGCCAGAAGCAAATGATCTCCTTTAGTCCTCAACAACAACAAC
 AGCAGCAGCAGTATATGGCCCAGTACTGGAGTGACACATTGAATCTGAGTCCAAGAGGAA
 GAATGATGATGATGATGAGCCAAGAAGCTGTTCAACCTTACATCGCAACGAAGCTGTACA
 GAGGAGTGAGACAACGTCATGGGGAAAATGGGTCGCAGAGATCCGTAAGCCACGAAGCA
 GGGCAGCTCTTTGGCTTGGTACCTTTGATACAGCTGAAGAAGCTGCCATGGCCTACGACC
 GCCAAGCCTTCAAATTACGAGGCCACAGCGCAACACTGAATTTCCCGGAGCATTTTGTGA
 ATAAGGAAAGCGAGCTGCATGATTCAAACCTCGTCGGATCAGAAAGAACCTGAAACGCCAC
 AGCCAAGCGAGGTTAACTTGGAGAGCAAGGAACCTACCGGTGATTGATGTTGGGAGAGAGG
 AAGGTATGGCTGAGGCATGGTACAATGCCATTACATCGGGATGGGGTCTGAAAGTCTCTC
 TTTGGGATGATTTGGATAGTTCTCATCAGTTTTTCATCAGAAAGCTCATCTTCTTCTCTC
 TCTCTTGTCTCTATGAGGCCTTTCTTTGAAAAAGTTTATAAACCACATTGTGTGTAGG
 TTATAGTTTGGGTTATGCTCATTGGCATTGGATGGAGGCAATTTTGTGATCTCCCAT
 TCCACCACATATCAGTCATTATATGTGTCTACCTTTTCTCTGTATTCTATCATTATCAT
 TGTTTTTATTATGTGTCTGTATGTGTTCCCTATTGCTACATACATAGATGTCTCTTTG
 TTCAAAAAAAAAAAAAAAAAAAAAA

>G1845 Amino Acid Sequence (domain in AA coordinates: 140-207)

MDFDEELNLCITKGNVDHVSFGGEASSTSPRSMKMKSPSRPKPYFQSSSSPYSLEAPPF
 SLDPTLQNNQQQLGSYVPVLEQRQDPTMQGQKQMISFSPQQQQQQQYMAQYWSDTLNL
 PRGRMMMSQEAQVQPYIATKLYRGVRQROWGKVAEIRKPRSRARLWLGTFTAEAAAM
 AYDRQAFKLRGHSATLNFPEHFVNKESELHDSNSSDQKEPETPQPSEVNLESKELPVIDV
 GREEGMAEAWYNAITSGWGPESPLWDDLSSSHQFSSESSSSSPLSCPMRPF*

>G1879 (3..917)

AAATGCCCTTAGAGGCTGTCTGATACCCGCAAGATCCATTCCGATATCTCTCCAATTGCA
 AAGATTTTATGTTCCACGACTTATACTCTCAAGAAGAGTTCGTAGCTCAAGATACGAAGA
 ACAACATTGATAAGTTAGGGCATGAACAGAGCTTTGTGGAACAAGGTAAGGAGGACGATC
 ATCAATGGCGAGACTATCATCAGTATCCTTTGTTGATCCCTTCGTGAGGAGAAGAGCTTG
 GTCTTACCGCCATTGATGTGGAGAGTCACTCTCTCCACAGCACCGGAGGAAGAGGAGGA
 GAACGAGAAACTGCAAGAACAAAGGAAGAGATCGAGAACCAGAGAACTGACTCACATCGCCG
 TCGAGAGAAATCGCCGAAACAGATGAACGAGTATCTGGCTGTGCTCCGTTCTCTAATGC
 CGTCTGTCGTATGCTCAAAGAGGAGATCAAGCGTCGATAGTAGGAGGAGCTATAAATACG
 TGAAGGAGTTAGAGCATATTTACAATCTATGGAGCCGAAGAGAACTAGGACTCATGATC
 CCAAAGGAGACAAGACTAGCACTAGCTCGTTAGTGGGTCCATTACAGATTTTTTTCAGCT
 TCCCACAATATCTACAAAGTCATCATCAGATGTACCGAAAGCTCATCTTCACCGGCGG
 AGATAGAGGTTACGGTGGCAGAAAGCCATGCGAACATCAAGATAATGACGAAGAAGAAAC
 CGAGGCAGCTTCTTAAGCTCATACTTCTTTACAAAGCCTAAGGCTCACTCTTCTTCATC
 TCAATGTCAACACTCTCCACAACCTCATCTCTACTCCATCAGCGTCAGGGTTGAAGAAG
 GAAGCCAACCTGAATACCGTGGACGACATTGCAACAGCTTTGAATCAAACCATAAGGAGGA
 TTCAAGAAGAGACATAATTCAGCAAATAGATTATAATTAACCTGTTTTATTTTATTTTA
 TTTTGAAATAACTGAAATCAGTTTTCTAATTTTTTTTTTTTTTCACTATTCCTCTAATCC
 TCCCTATGTAAGTTGATTTTGTCTCTGTAATGAATCAATGGTCATAAAGATCTGAAC
 AAAAAAATTGAATAAAAGAAAATGGTT

>G1879 Amino Acid Sequence (domain in AA coordinates: 107-176)

MPLEAVVYPQDPFGYLSNCKDFMFHDLYSQEEFVAQDTKNNIDKLGHEQSFEQKEDDH
 QWRDYHQYPLLPISLGEELGLTAIDVESHPPPHRRRKRRTNRCKNKEIENQRMTHIAV
 ERNRRKQMNEYLAVALRSLMPSSYAQRGDQASIVGGAINYVKELEHILQSMERKTRTHDP
 KGDKTSTSSLVGPFTDFFSFPQYSTKSSSDVPESSSSPAEIEVTVAESHANIKIMTKKKP
 RQLLKLITSLQSLRLTLHLNVTTLHNSILYSISVRVEEGSQLNTVDDIATALNQITIRRI

QEET*

>G1888 (1..729)

ATGAAGATTTGGTGTGCTGTTTGTGATAAAGAAGAAGCTTCGGTGTTTTGTGTGCGGAT
GAAGCAGCTCTTTGTAATGGTTGCGATCGCCATGTTTCATTTTCGCCAATAAACTAGCCGGG
AAACATCTCCGGTTCTCTCTCACTTCTCCTACTTTCAAAGATGCTCCTCTTTGTGATATT
TGCGGGGAGAGGCGTGCATTATTATTTTGCCAAGAAGACAGAGCAATACTATGCAGAGAA
TGTGACATTCCAATACATCAAGCTAATGAGCACACTAAGAAACACAATAGATTCTCTCCTT
ACCGGCGTTAAGATCTCTGCTCCCGTCAGCCTACCCAAGAGCCTCCAATTCCAATCTCT
GCTGCTGCATTTGGTCGAGCCAAAACCCGACCAAAATCAGTATCGAGCGAGGTCCCAGC
TCGGCCTCCAATGAGGTATTTACGAGCTCTTCTTCGACGACCACGAGCAATTGCTATTAT
GGGATAGAAGAAAACCTACCATCACGTGAGCGATTCCGGGTCGGGATCGGGTTGTACAGGT
AGTATATCCGAGTATTTGATGGAGACATTACCGGGTTGGAGAGTGGAGGATTTGCTTGAA
CACCTTCTTGTGTCTCCTATGAGGATAACATTATTACTAATAACAATAACAGTGAGTCT
TATAGGGTTTATGATGGTTCCTCACAAATTCATCATCAAGGGTTTGGGATCACAAACCC
TTCTCTTGA

>G1888 Amino Acid Sequence (domain in aa coordinates: 5-50)

MKIWCAVCDKEEASVFCCAEALCNGCDRHHVFANKLAGKHLRFSLTSPTFKDAPLCDI
CGERRALLFCQEDRAILCRECDIPIHQANEHTKKHNRFLLTGVKISASPSAYPRASNSNS
AAAFGRAKTRPKSVSSEVPSSASNEVFTSSSSTTNSNCYGIEENYHHVSDSGSGSGCTG
SISEYLMETLPGRWRVEDLLEHPSCVSYEDNIITNNNSES YRVYDGSSQFHHQGFWDHKP
FS*

>G189 (34..987)

CCACAACCTCTCTCCTTGTAGAGAGAGAGATTTTATGGCGGTGGAGCTCATGACTCGGAAT
TACATCTCCGGCGTCGGAGCTGATAGCTTCGCGGTTCAAGAAGCAGCTGCTTCAGGACTC
AAAAGTATCGAAAAATTCATCGGTTTAATGTCTCGTGATAGCTTAACTCTGATCAGCCA
TCTTCTTCTTCCGCTCCGCTCCGCTCCGCTCCGCGCCGAGATCTTGAATCAGCTCGTAAC
ACAACGGCGGACGCGGCTGTTTCAAAGTTTAAAAGAGTCATATCTCTCTTAGATCGAACT
CGAACCGGACACGCCCGGTTTAGACGTGCTCCGGTTCATGTTATTTCTCCGGTCTCTTTA
CAAGAAGAACCAAAAACGACGCCGTTTCAAGTCTCCTCTCTCCTCCCGCCGAAATGATC
CGAAAAGGTTCTGTTTCTTCATCGATGAAAACGATTGATTTCTCATCTCTCTCCTCTGTA
ACAACCGGAATCAGACAACCAAGAAGATTTCATCATCATCAACGTCCCTCTGAAACGGCG
CCGTTTGCGTCTCAAACCTCAAAGCCTCTCCACGACGGTCTCGTCTTTCTCAAATCAACA
AAGAGAAAATGTAACCTGAGAATCTTCTCACCGGAAATGCGCTTCCGCTTCTTCTCCTCC
GGTCGTGTGTCATTGTCTCGAAGAAAAGATAAAACAGAGGAGAATAATTAGGGTCCG
GCGATAAGTGCAAAAATGTCCGATGTACCACCGGACGATTATTCATGGAGGAAATACGGA
CAAAAACCAATTAAAGGATCTCCACATCCAAGAGGATATTATAAGTGATAGCGTAAGA
GGTTGTCCAGCACGTAAACATGTTGAGAGAGCAGCTGATGATTCTGTCATGTTGATTGTT
ACTTATGAAGGAGATCATAATCATTCTCTCTCCGCGCTGATCTCGCCGGAGCCCGCGTT
GCTGATCTTATTTTGGAAATCGTCTTGAAAAGAACAATCTTTATTTAAGGCTTTTATAAT
ATAAATTTAGATCCTTACTTAGTGAAGTACTCAAACATGAATGAAATCAATGTAATCAA
AATCAAAAAGCTTTTGCTAAAAA

>G189 Amino Acid Sequence (domain in AA coordinates: 240-297)

MAVELMTRNYISGVGADSFVQEEAAASGLKSIENFIGLMSRDSFNSDQPSSSASASASA
AADLESARNTTADA AVSKFRVISLLDRTRTG HARFRAPVHVISP VLLQEEPKTTPFQS
PLPPPPQMIRKGSFSSSMKTIDFSSLSVTTESDNQKKIHHHQRPS ETAPFASQTQSLST
TVSSFSKSTKRKCNSENLLTGKASASSSGRCHCSKKRKIKQRRIRVPAISAKMSDVPP
DDYSWRKYGQKPIKGSPPHPRGYKYCSSVRGCPARKHVERAADDSSMLIVTYEGDHNHSL
AADLAGAAVADLILESS*

>G1939 (92..844)

AATCATTAGTCTCTTCTCTCTCTCTCTCACAGAGAGAGTAATCACAAGCCAAGTGAGA
AAAAGAAAACCTAAACCCAGATCGAAAACCATGTCTATTAAACAACAACAACAACA
CAACAATAACAACGATGGTCTTATGATCTCATCAAACGGAGCTTTAATCGAACAACAACC
ATCAGTCGTTGTGAAGAAACCACCGCGAAAGATCGACATAGCAAAGTCGATGGAAGAGG
GAGAAGAATCCGATGCCGATTATATGTGCTGCTCGTGTTTTTCAGCTAACGAGAGAGCT
TGGTCATAAGTCAGATGGCCAAACAATGAATGGTTACTTCGTCAAGCAGAGCCTTCTAT
TATAGCTGCAACAGGAACCTGGTACAACCTCCAGCGAGTTTCTCAACTGCTTCTGTCTCTAT
CCGTGGAGCCACCAATTCTACTTCTTTAGATCATAAACCCACTTCTTTACTTGGTGGTAC

GTCACCGTTTATACTTGGGAAACGTGTTAGAGCTGATGAGGATAGTAATAATAGTCATAA
TCATAGTTCTGTTGGTAAAGATGAGACCTTTACGACAACACCAGCTGGGTTTTGGGCTGT
TCCGGCGAGGCCGGATTTTGGACAAGTTTGGAGTTTTGCTGGAGCTCCAACAAGAGATGTT
TTTACAACAACAACATCATCATCAGCAACCATTGTTTGTTCATCAGCAACAGCAACAACA
AGCTGCAATGGGTGAAGCTTCTGCTGCTAGAGTTGGGAATTATCTTCCGGGTCTATCTTAA
TTTGCTTGCTTCTTTATCCGGTGGATCTCCCGGTCCGATCGAAGAGAGGAAGATCCACG
TTAATGGTTTAAAGCCCTTTTAGGTTTGAAGGCAAAATTTGGTATATATATTTATTATCTT
CTCTTCTCTATTGTTGTCTATTGTTTCTCTATGTGTGTGTTTTAGTGTGTTAGAGATTGA
TTTGGTTTCAGAATCTCTGCAAGTGATTTGAGAGTTTTCGTTAGCTTTAAGTAAGTTAAA
GACGGTTGTTTTTGATTAGGGTTAAATTAGGGTTTAAAGATCTGTTGTTTTTTTGGAGGG
AGATCGATTTCTTATCGGATCCAAGATTACTTTTAGGAAAAAAGGGAAAATTTTCAGAAAC
CACGGTGGTTTCTTTTCTCTTTTTTTTTTTTTT

>G1939 Amino Acid Sequence (domain in AA coordinates: 40-102)
MSINNNNNNNNNNDGLMISSNGALIEQQPSVVVKPPAKDRHSKVDGRGRIRMPIICA
ARVFQLTRELGHKSDGQTIEWLLRQAEPsIIAATGTGTPASFSTASVSIRGATNSTSLD
HKPTSLGGTSPFILKRVRAEDSDNNSHNHSSVGKDETFTTTPAGFWAVPARPDFGQVW
SFAGAPQEMFLQQQHHHQPLFVHQQQQQAAMGEASAARVGNYPGLHNLASLSGGSP
GSDRREEDPR*

>G194 (192..1205)

TCTTTCTCTCTCTCTATCTCTCTCTTTGAACCCTAAAACTCTTTCTTTACAAGGATT
GATCTTTTTGTATTTTGTATTTGACATTGCTTTGTGTTTCGATCTCTGTTTTGATGCGA
TTTCTCTGTTTTTAAAGCCATTGATAGATTGTTTCCGGTAAAGCTCAGCGAGAGAAGAA
GAAGAACAACAATGGAGTTTACAGATTTCTCAAAGACGAGTTTTTACTACCCGTCTGTCAC
AAAGCGTTTGGGATTTTCGGAGATTTAGCGGCGGCGAGAGGCATTCTTAGGGTTCATGG
AGTTATTAAAGTTCTCAGCAGCATCAAGACTTTGCTACTGTTTCTCTCATTCTCTCTTC
TCCAAACGTCTCAACCGCAAACGCAAACGCAACCATCGGCGAAGCTGTCTTCAAGTATCA
TTCAAGCTCCACCGTCAGAGCAATTAGTGACGTCAAAGGTGGAGTCTTGTGTTTCGGATC
ATTTGTTGATAAACCCACCGGCGACTCCTAACTCGTCATCGATTTCGTCTGCTTCAAGCG
AGGCTCTAAATGAAGAGAAACCGAAAAACAGAAGACAATGAAGAAGAAGGAGGTGAAGATC
AACAAGAGAAGAGTCATACTAAGAAACAGTTGAAAGCAAAGAAGAATAATCAGAAGAGAC
AGAGAGAGGCAAGAGTCGCATTCTATGACAAAGAGTGAAGTTGATCATCTCGAAGATGGTT
ATCGCTGGCGAAAAATATGGTCAAAAAGCTGTCAAAAACAGTCCTTTTCCAGGAGTTACT
ACCGTTGCACAACGGCTTTCATGTAACGTGAAGAAGAGAGTGGAGAGATCATTTCAGAGATC
CAAGCACTGTGGTTACAACCTACGAAGGTCAACACACTCACATTAGTCCACTCACGTCTC
GTCCTATTTCCACTGGAGGTTTCTTCGGATCGTCAGGAGCTGCTTCGAGTCTCGGTAATG
GTTGCTTTGGGTTTCTTATGATGGCTCCACGTTAATCTCTCCTCAGTTCCAACAGCTTG
TCCAATACCATCACCAACAGCAGCAACAAGAATCATGTCTTGTGTTTGGAGGAGTCAACG
AGTACCTTAATAGCCACGCTAATGAGTATGGTGATGATAATCGTGTGAAGAAGAGTCGAG
TTTTGGTTAAAGATAATGGACTTCTGCAAGATGTTGTTCCGTCTCATATGTTGAAGGAAG
AGTAGTAGTATATATATAGTCTTATAGTTTTAATCTAGTTTTTTTTTGTATAATTGTCTA
AAAGAAACGGATCTTTTGTTCGATGAAGAAGATGTTTCTTATGGTTCTGAAATCGTAA
GGTAATGATGATTGTACCAAGCCGAGAAAGTACTTGTGATTTTACCATTGAATCACTAT
AAATGTAATTTTTTATTACTGTGAAAAAAAAAAAAAAAAA

>G194 Amino Acid Sequence (domain in AA coordinates: 174-230)
MEFTDFSKTSFYYPSSQSVWDFGLAAAERHSLGFMELLSSQQHQDFATVSPHSFLLQTS
QPQTQTQPSAKLSSSIIQAPPSEQLVTSKVESLCSHLLINPPATPNSSSISSEALN
EEKPKTEDNEEEGGEDQEQESHKKQLKAKKNNQKRQREARVAFMTKSEVDHLEDGYRWR
KYGQKAVKNSFPFRSYRCTTASCNVKKRVERSFRDPSTVVTTYEQHTHISPLTSRPIS
TGGFFGSSGAASSLNGCGFGFPIDGSTLISPQFQQLVQYHHQQQQELMSCFGGVNEYLN
SHANEYGDDNRVKRSRLVKDNGLLQDVVPSHMLKEE*

>G1943 (137..1858)

ACATTTGTTTCTAATCTCAGACATAAATAATTTTTGTTCGGACTTCAAAACCAACGATG
ATTATATCATTCACATTCTTTCTTCTACTTCTTCTCTCTCTTCTCTCTTCTCTCTCT
AGAAAATCCATCTATCATGGGTGAAGATGATATAGTGAGCTCTTATGGAAGAGTGGCCA
AGTCGTTAGAACCAGTCAAACAGAGACCCTCCTCCAATACACCACCATCTCTCTCTCC
ACCACCCATTCTTCGTGGTAGCGGAAGCGGCAACGGAGAAGAAAATGCCCCGCTTCCACT
TCCACAGCCTTCACCTCCCTCCATCATCAGAATCTTTTCATTCTGGAAGACGAAATGTC

TTCTTGGCTTCACCATTCTACCCCGGCGTTACGTCCACCCCGGCTTCTTCTGTCTCCCT
GCCACCACCACCCAATGCTCCGCGTGAAGATGATATAGTGGAGCTTTTATGGCAAAGCGG
CCAAGTAGTTGGAACCAACCAAACACATAGACAATCCTACGATCCTCCTCCCATTTCTCCG
CGGCAGCGGAAGTGGCAGAGGAGAAGAAAATGCTCCCTTTTACAACCTCCGCTCACCT
GCATCAGCAAAATCTCTTCATTCAAGAAGGCGAAATGTATTTCGTGGCTACACCATTCTTA
CCGCCAAAATATTTCTGTCTCAGAACTTCTCAACTCCACTCCGGCTACTCACCCGCAAAG
TTCCATCTCTCTGGCACCAGTCAGACTATCGCCACGAGAAGGGCGGAAAACCTTTATGAA
CTTCTCGTGGCTAAGAGGGAACATATTTACCGGCGGTAGAGTTGATGAAGCTGGACCGTC
GTTTTCGGTGGTAAGAGAATCGATGCAGGTAGGCTCGAACACGACCCCCCTTCTCTTC
TGCCACTGAATCATGTGTAATACCAGCTACAGAGGGCACCAGGAGTCGAGTGTGGGAAC
TTTGGCAGCTCATGATCTTGGTCGGAAGGGAAGGCGGTGGCGGTTGAGGCGGCCGAAC
ACCATCTTCAGGAGTGTGCAAGGCCGAAACAGAGCCGTTTCAGATACAACCAGCAACGGA
GTGCAAGCTAAAAGCGAGAGAAGAAACCCATGGAAC'TGAAGAAGCTCGTGGTTCAACGTC
TAGAAAGAGATCACGAATGCAGAAATGCATAACCTCGCCGAAAGGAGAAGGAGAGAAAA
GATCAACGAGAAGATGAAGACTCTGCAACAACCTCATTCTCGCTGCAACAAGGTTGAATC
TGATTCTGTTTTCTACTCTGATCAGTCTACTAAAGTTTCAACGCTGGATGATGCTATCGAG
TACGTCAAATCGTTACAGAGCCAAATACAAGTATGCTCTTCAAAACAGAATGTGTTTAA
ACCAATGGTTCAACATGGAAGAGTTCATATGTATCTAGTTTTGTTGAGATGATGTCGAC
GGGACAGGGTATGATGTCGCAATGATGAATGCCGGAATACGCAACAGTTTCATGCCCA
TATGGCCATGGATATGAACCGACCTCCTCCATTCATACCTTTCCCGGCACATCTTTCC
TATGCCGGCTCAAATGGCAGGTGTAGGTCCATCATATCCAGCACCGCGCTACCTTTTCC
CAACATTCAGACCTTTGACCCATCCAGAGTCCGTTTACCAAGCCCGCAGCCTAACCCGGT
GTCGAACCAGCCTCAGTTTCCGGCTTACATGAATCCCTATAGCCAGTTTGCTGGTCCCA
CCAGTTGCAACAACCTCCTCCTCCTCCATTTCAGGGTCAAACAACATCACAACTGAGTTC
CGGGCAGGCAAGTAGTAGCAAGGAACCTGAGGATCAGGAGAACCACCAACAGCTTAGTT
AAAGTGTGGAGCTGAAACGGATCAGTTCTTCAAGCAAATTACAACCTTTGAAGATAAACCA
GAGTTGTAACATGTAGATTTTGTCTGTTAAGTTTAATGTAAGTACTTTTAGTTAATGGG
AAAGATACTGACAGGTTGCAAGGTGGTCAGTATTTGTGCATCACGCTTAAGATTCTCGA
TGTGGCCAGTATCTCCCTTTTCTAGCATGTGAGGTCCCTACTCTCTGGTTCTACGGAGAC
CAAATGTTTCGACTGATTAACACACAATGACTTACCAAAAGTACACGGGCCCCATCCTCG
TCTTTATGTTCCAAGTGCAGCTGTTTGTATTGTAAGCATTTTTCTTATAATAATAAA
ACAGCTCTATCTTCGTTAAAAAAA

>G1943 Amino Acid Sequence (domain in AA coordinates: 335-406)

MGEDDIVELLWKSGQVVRTSQTRPSSNTPPSLPPPPILRGSGSGNGEENAPLPLPQSP
PLHHQNLFILEDEMSSWLHSHHPGVTSTPASSVSLPPPPNAPREDDIVELLWQSGQVVG
NQTHRQSYDPPPILRGSGSGRGEENAPLSQPPPHLHQNLFIQEGEMYSWLHHSYRQNYF
CSELLNSTPATHPQSSISLAPRQTIATRRANFMNFSWLRGNIFTGGRVDEAGPSFSVVR
ESMQVGSNTTPSSSATESCVIPATEGTASRVSGTLAAHDLGRKGKAVAVEAAGTPSSGV
CKAETEPVQIQPATESKLKAREETHGTEEARGSTSRKRSRTAEMHNLAEERRRREKINEKM
KTLQQLIPRCNKVESDSVSTLISLLKFQRWMMLSSTSNRYRAKYKYALQNRMCFKPMVQH
GKSSYVSSFVEMMSTGQMMSPMMNAGNTQQFMPHMAADMNRPPPFIPFGTSFPMQAQ
AGVGPSYPAPRYPPFNIQTFDPSRVRLPSPQPNPVSNPQPPAYMNPYSQFAGPHQLQPP
PPPPFQGQTTSQLSSGQASSSKEPEDQENQPTA*

>G21 (79..966)

TGTGGAGGAATATTAATACAGCCCACTTCACATCTATTTTGTGCAACCATCTCTCTAAA
GCTTCTTCTCTCATAACAATGGCAAGACAAATCAACATAGAGAGTAGTGTCTCAAGTT
ACCTTTATCTCCTCCGCCATCCCCGCCGTATCTTCTCCTCCTCCATCACCGCTTCCGCC
TCATTGTCTCTTACCTACTACATCTTCTCTTCTCGTCATCAACAAATTCTAACTTC
ATTGAGGAAGACAACCTATAAAGAAAAGCATCTCGAAGATCATTGTCTATCGTTAGTCTCC
GTTGAAGACGATGATGATCAAAACGGTGGAGGTGGGAAACGGCGAAAGACCAACGGTGGA
GATAACATCCGACGTATAGAGGAGTGAGGATGAGGAGTTGGGGAAAATGGGTGTGCGAG
ATTAGAGAGCCGAGAAAGAAATCAAGAATCTGGCTCGGGACTTATCCAACGGCTGAGATG
GCAGCTCGAGCTCATGACGTAGCGGCTTTAGCCATTAAAGGTACAACGGCTTACCTCAAT
TTTCCCAAGTTAGCCGGCGAGCTTCTCGTCCAGTCACAAATCTCTCTAAAGACATTCAA
GCCGCCGCCTCTTTAGCGGGCGGTTAACTGGCAAGATTCCGGTCAACGATGTGAGTAATTCT
GAAGTGGCTGAAATAGTTGAAGCCGAGCCGAGTCGAGCCGTGGTGGCTCAGTTGTTTTCT
TCGGACACAAGCACGACGACGACTCAGAGTCAAGAGTATTCGGAAGCTTCGTGTGCT

TCGACTTCGGCGTGTACGGACAAAGACAGTGAGGAAGAGAAGCTGTTTGATTGCGCGAT
 TTGTTTACCGATGAGAATGAGATGATGATACGAAACGATGCGTTTTGCTACTACTCGTCC
 ACGTGGCAGCTTTGTGGAGCCGATGCTGGGTTTCGGCTTGAAGAGCCGTTTTTCTATCT
 GAATGACTAAAGTACCCCTCTCGAGAGAGCTCTCACTAACACT

>G21 Amino Acid Sequence (domain in AA coordinates: 97-164)
 MARQINIESSVSQVTFISSAIPAVSSSSSITASASLSSSPTSSSSSSSTNSNFIEEDNS
 KRKASRRSLSSLVSVEDDDDDQNGGGKRRKTNGGDKHPTYRGVVRMRSWGKWWSEIREPRK
 KSRIWLGTYPTEAEMAARAHDAALAIKGTAYLNFPLAGELPRPVTNSPKDIQAAASLA
 AVNWQDSVNDVSNSEVAEIVEAEPSRAVVAQLFSSDTSTTTTTSQSEYSEASCSTACT
 DKDSEEEKLFDLPDLFTDENEMMIRNDAFCYSSSTWQLCGADAGFRLEEPFFLSE*

>G2132 (42..1031)

ATTCTGTACTTAGTACCGGAGTTTAGTCGGAGAGAGACAATGATCAGTTTCAGAGAAG
 AGAACATCGATCTCAACTTGATTAAACAATTAGTGTATCTGTAATGATCCAGACGCCA
 CCGATTCTCTAGCGACGATGAATCTATCTCCGGCAATAATCTCGCCGTCAGATCAAAC
 CAAAACCACCGAAACGTTACGTCTCAAAGATCTGTGTCCGACGCTGATCAAAGGTATG
 AGAACGTTTCGAATTCTACAGGGAATAAGCAGCCGGAACCGGAAACGTCGTCGGGTT
 TCAAAGGCGTACGACGAGGCCGTGGGGGAAATTTGCGGCGGAGATAAGAAAATCCGTTTG
 AGAAGAAGAGAAAAGTGGCTTGAACGTTTCTACTGAAGAAGAAGCAGCAGAAGCTTACC
 AAAAGAGTAAAAGAGAGTTTGATGAACGATTGGGTTTAGTTAAACAGGAAAAAGACCTAG
 TAGATTTGACCAAGCCGTGCGGTGTACGTAAACCAGAAGAGAAGGAAGTTACTGAGAAGT
 CGAATTGCAAAAAGGTAAATAAGAGAATTGTTACTGATCAGAAGCCATTGTTGTGGTT
 ATAACGCTGATCATGAAGAAGAGGGAGTGATTAGTAAATGTTGGAAGATCCGTTGATGA
 CATCGTCAATTGCTGATATTTTGGTGATTGCGCTGTTGAAGCAAATGATATTTGGGTGG
 ATTACAATTCAGTGAATTTATTTCCATTGTAGATGATTTCAAGTTTGATTTTGTGGAGA
 ATGATAGAGTAGGAAAGGAGAGAAAACATTTGGATTAAAGATTGGGGATCACACTAAAGTTA
 ATCAACATGCCAAAATCGTATCGACCAATGGGGACTTATTCGTCGATGATTTACTTGATT
 TTGATCCGTTGATAGATGATTTTAAGTTAGAAGATTTTCCTATGGATGATCTTGGATTAT
 TAGGAGATCCAGAGGATGATGATTTTAGTTGGTTTAATGGTACTACTGATTGGATCGATA
 AGTTTTTATGAATACTTTCTTGACACGCCAACGGTATTAGTAC

>G2132 Amino Acid Sequence (domain in AA coordinates: TBD)
 MISFREENIDLNLIKTISVICNDPPATDSSSDDESISGNPNRRQIKPKPPKRYVSKICVP
 TLIKRYENVSNSTGNKAAGNRKTSSGFKGVRRRPWGKFAAEIRNPFKKRWLGTFTPTEE
 EAAEAYQKSKREFDERLGLVKQEKDLVDLTKPCGVVRKPEBEKTEKSNCKKVNKRIVTDQ
 KPFGCGYNADHEEERGVIKMLEDPLMTSSSIADIFGDSAVEANDIWVDYNSVEFISIVDDF
 KFDFVENDRVGKEKTFGFKIGDHTKVNQHAKIVSTNGDLFVDDLLDFDPLIDDFKLEDFP
 MDDLGLLGDPEDDDFSWFNGTTDWDKFL*

>G2145 (1..777)

ATGGACGTTTTTGTGATGGTGAATTGGAGTCTCTCTTGGGGATGTTCAACTTTGATCAA
 TGTTTCATCATCTAAAGAGGAGAGACCGCGAGACGAGTTGCTTGGCCTCTCTAGCCTTTAC
 AATGGTCATCTTCATCAACATCAACACCATAACAATGTCCTTATCTTCTGATCATCATGCT
 TTCTTGCTCCCTGATATGTTCCCATTTGGTGCAATGCCGGGAGGAAATCTTCCGGCCATG
 CTTGATTCTTGGGATCAAAGTCATCACCTCCAAGAAACGTCCTTCTTAAAGAGGAACTA
 CTTGACGTGGAGAATCTATGCAAACTAACTCTAACTGTGACGTCACAAGACAAGAGCTT
 GCGAAATCCAAGAAAAAACAGAGGGTAAGCTCGGAAAGCAATACAGTTGACGAGAGCAAC
 ACTAATTGGGTAGATGGTCAGAGTTTAAGCAACAGTTTCAAGATGATGAGAAAGCTTCGGTC
 ACAAGTGTTAAAGGCAAACTAGAGCCACCAAGGGACAGCCACTGATCCTCAAAGCCTT
 TATGCTCGGAAACGAAGAGAGAAGATTAACGAAAGGCTCAAGACACTACAAAACCTTGTC
 CCAAACGGGACAAAAGTCGATATAAGCACGATGCTTGAAGAAGCGGTCCATTACGTGAAG
 TTCTTGACGCTTCAGATTAAGTTGTTGAGCTCGGATGATCTATGGATGTACGCACCATTG
 GCTTACAACGGCTTGGACATGGGGTTCCATCACAACCTTTGTCTCGGCTTATGTGA

>G2145 Amino Acid Sequence (domain in AA coordinates: 166-243)
 MDVFVDGELESLLGMFNFQDCSSSKEERPRDELLGLSSLYNGHLHQHNNVSSDHHA
 FLLPDMFPFGAMPGGNLPAMLDSDQSHHLQETSSLKRKLLDVENLCKTNSNCDVTRQEL
 AKSKKKQRVSSESNTVDESNTNWVDQSLSNSSDDEKASVTSVKGKTRATKGTATDPQSL
 YARKRREKINERLKLQNLVPNGTKVDISTMLEEAVHYVKFLQLQIKLLSSDDLWMYAPL
 AYNGLDMGFHNNLLSRLM*

>G23 (22..732)

TATCAAACGAGAGTACAAAAGATGACGTCACTCAACAGCTCTGCATCACCAACATCATCG
TCATCAGACCAATCTGATGCAACTACTACAACAAGCACCCACTTGTCTGAAGAAGAAGCT
CCACCCAGAAAACAACAACAAGAAAGAGAAGGAGAGATTCTTCTTGCTTCTTCATCT
TCTTCAATGCAACATCCTGTTTACAGAGGTGTGCGGATGAGAAGTTGGGGCAAATGGGTC
TCCGAGATCCGACAACCTCGTAAGAAAACCTCGTATTTGGCTCGGCACCTTTTGTACCCGCT
GATATGGCTGCTCGTGCTCACGACGTGCTGCTCTCACCATCAAAGGCTCCTCCGCCGCTC
TTAAATTTCCCTGAGCTTGCTTCTCTCTTCCCTCGTCCGGCGTCATCATCGCCGCATGAT
ATCCAGACAGCCGCCGAGAAGCCGCCGCGCATGGTGGTTCGAAGAAAACTGTTAGAGAAG
GATGAGGCTCCGGAGGCCCCACCTTCGTCCGAATCTTCTTACGTGGCGGCGGAGTCAGAG
GATGAGGAGAGGTTGGAGAAAATTGTGGAGCTGCCTAACATTGAAGAAGGAAGTTATGAC
GAGAGTGTGACATCACGTGCTGATCTGGCTTATTCTGAGCCGTTGATTGTTGGGTGTAT
CCTCCGGTTATGGATTTTTATGAAGAAATATCGGAGTTTAAATTCGTGGAATTGTGGAGC
TTTAATCACTAATTAAGTTAGGAAAGTGCATTATATTGCAATATTGCATCATAGATAACA
TTTGTATTTCTTTTCTTTTGTACGGATACGTAGCATATGCTACTATACTAGGGCTAGTG
TACCAAATATTGTAAAATATACTTATTAATATTTATGTAAATGTGTAATATATATAACAT
ACAATTATTGTAAGTTTGAAATTGAAAATATCGTTACGCAATGTTCTTGTAACAAAAA
AAAAA

>G23 Amino Acid Sequence (domain in AA coordinates: 61-117)
MTSLNSSASPTSSSSDQSDATTTTSTHLSSEEEAPPRNNNTRKRRRSDSSSSSSMQHPV
YRGVRRMSWGKWWSEIRQPRKKTRIWLGTFTADMAARAHVDVAALTIKSSAVLNFPELA
SLFPRPASSSPHDIQTAAABAAAMVVEEKLLEKDEAPEAPPSSESSYVAAESEDEERLEK
IVELPNIEEGSYDESVTSRADLAYSEPFDCWVYPPVMDFYEEISEFNFVELWSFNH*

>G2313 (104..724)

CGTCGACACAATCGCTCTTCCGTAACATATTCCACAAAACGATCTTCTTGTCTTGAAT
TTTTAGCCATCTCTTTTTTTTTTCTCATTTTCTCGGATACATATGGCTTCGAGTCCACG
CTGGACGGAGGACGACAACAGGCGTTTTAAGTCAGCTCTGTGCGCAATCCCTCCGGATAA
CAAGCGTTTGGTGAATGTGCGCCAGCATCTGCCGAAACCTTTGGAGGAGGTGAAGTACTA
CTACGAAAAGTTGGTCAACGATGTTTATCTGCCGAAACCTTTAGAGAATGTCACCCAGCA
TCTGCAGAAACCTATGGAAATGGAGGAGATGAAGTACATGTACGAAAAGATGGCCAACGA
TGTTAATCAGATGCCCGAGTACGTACCACTGGCGGAATCGAGTCAGTCCAAACGCAGGAA
GAAGGATACGCCAAATCCTTGGACAGAAGAGGAACACAGATTGTTTCTGCAAGGATTGAA
AAAGTATGGGGAAGGAGCTTCGACGTTGACATCAACGAATTTTGTGAAGACAAAGACTCC
ACGGCAAGTGTCAAGCCATGCACAGTATTACAAAAGGCAAAAATCGGACAATAAGAAGGA
GAAACGCCGGAGTATTTTGACATAACTTTGGAGTCTACCGAGGGCAATCCAGATTCTGG
AAATCAGAACCCTCCGGATGATGATGATCCGTCCCAAGGTCAAGGCACTTGTCTTGAGT
TTAGATGTTTGAAGATAGAAGATGGTGTGAAAGC

>G2313 Amino Acid Sequence (domain in AA coordinates: TBD)
MASSPRWTEDDNRRFKSALSQFPDPNKRLLVNVAQHLPKPLEEVKYYYEKLVDVYLPKPL
ENVTOHLQKPMEMEEMKMYEKMANDVNQMPEYVPLAESSQSKRRKDTNPWTEEEHRL
FLQGLKKYEGASTLTSNFKTKTPROVSSHAQYYKQKSDNKKERRSIFDITLESTE
GNPDSGNQNPDDDDPSQGQGTCLGV*

>G2344 (1..573)

ATGACTTCTTCAATCCATGAGCTTTCTGATAACATTGGAAGTCATGAGAAGCAAGAACAG
AGAGATTCTCATTTCCAACCAATCCCTTCTGCAAGAAATTATGAATCAATTGTTACA
AGTTTAGTCTACTCAGACCCGGGGACTACAAATCCATGGCACCTGGACAATATCCATAT
CCAGATCCTTACTACAGAAGCATATTTGCACCGCCTCCACAACCGTATACCGGGGTACAT
CTACAGTTGATGGGAGTGCAGCAACAAGGCGTTCTTTTACCATCTGATGCAGTCGAGGAA
CCTGTTTTTGTAAEGCAAAGCAATACCACGGTATACTAAGGCGCAGACAATCAAGAGCA
AGACTTGAGTCTCAGAATAAAGTCATCAAGTCACGTAAGCCGTATTTGCATGAATCTCGG
CATTTGCATGCGATAAGACGACCAAGAGGATGTGGCGGGCGGTTTCTAAATGCCAAGAAG
GAGGATGAGCATCACGAAGACAGTAGTCATGAAGAAAAATCCAACCTTAGCGCTGGTAAA
TCCGCCATGGCTGCTTCTAGTGGTACATCTTGA

>G2344 Amino Acid Sequence (domain in AA coordinates: TBD)
MTSSIHELSDNIGSHEKQEQRDSHFQPPIPSARNYESIVTSLVYSDPGTTNSMAPGQYPY
PDPYYSIFAPPPQPYTGVHLQLMGVQQGVPLPSDAVEEPVFNQKQYHGILRRRQSRA
RLESQNKVIKSRKPYLHESRHLHAIRRPGRGGRFLNAKKEDEHHEDSSHEEKSNLISAGK
SAMAASSGTS*

TCCTCACTCTCTCTCTTTTCTCTAACCATAAAATCTCTTTGATCTCTTTCTCTGTGTTT
TGATAATGGAAAATGTTGGTGTTGGGATGCCGTTTACGATTTAGGGCAAACAAGGGTTT
ACCCACTCTTGCTGATTTCCACGATTTACCGCGGAGAGGTATCCGGTAGGGTTCATGG
ATTTACTGGGTGTTTCATCGTCATACACCCACCCTACGCCGTGATGCATTTTCCGACCA
CACCTAACTCGTCTCGAGCGAAGCTGTGAATGGAGATGACGAAGAAGAAGATGGAG
AAGAACAGCAGCATAAGACAAAGAAGCGGTTTAAATTCACTAAAATGAGTAGAAAGCAGA
CGAAGAAGAAGGTGCCAAAAGTGTCAATTCATCACGAGGAGTGAGGTTCTTCATCTAGATG
ATGGTTATAAGTGGAGAAAATACGGTCAAAAACCTGTCAAAGACAGCCCTTTTCCAAGAA
ATTATTACCGTTGCAACAACAATTGGTGTGACGTGAAGAAGAGAGTAGAGAGATCATTTCA

GTGATCCAAGCAGTGTAATCACCACCTTACGAAGGTCAACATACTCATCCTCGTCCACTAC
TCATCATGCCCAAAGAAGGCAGCTCTCCATCCAATGGCTCAGCTTCTAGGGCCACATTG
GCCTCCCTACACTCCCTCCTCAGCTTTTAGATTACAACAACCAACAACAAGCGCCGT
CTTCTTTTGAACCGAGTACATTAAACAGGCAAGAAAAAGGAATTAATCATGATGATGATG
ACGATCATGTTGTGAAGAAGAGTCCGAACCTCGGGATCTGCTGGATGGAGCTGGTTTAGTCA
AAGATCATGGCCTTCTTCAGGATGTTGTTCCCTCTCATATCATTAAGGAAGAGTATTAGT
TAATCGCATAATTATGTAGCTAGCTAGCTAG

>G2517 Amino Acid Sequence (domain in AA coordinates: TBD)
MENVGVMGPFYDLGQTRVYPLLSDFHDLAERYPVGFMDLLGVHRHTPTHTPLMHFPPTTP
NSSSSSEAVNGDDEEEEDGEEQHKTKKRFKFTKMSRKQTKKKVPKVSFITRSEVLHLLDDG
YKWRKYGQKPKVDSPPFRNYRCTTTWCDVKKRVERSFSDPSSVITTYEGQHTHPRPLLI
MPKEGSSPSNGSASRAHIGLPTLPQLLDYNNQQQAPSSFGTEYINRQEKGINHDDDDDD
HVVKKSRTDRLDGLVKGHLLQDVVPSHIIKEY*

>G2521 (103..768)
ATTCTCCACAATTTTCATAACTTTCTTCCGCTCAACTTCAGATAAATTCGGATTCTGTAGC
TCTTTCAATACGACTGCGGAGATCAGAGCCAATTATTTGGTTATGGCGTCTCTGATCTCA
GATATTGAACCGCCGACGAGTACTACTTCAGATCTCGTTCGGAGAAAGAAGAGATCCTCT
GCTTCATCCGCCGATCGTCTCGTTCAAGCGCATCTTCCGCTCTCCGGTGAGATTCACGCG
CGATGGCGATCGGAGAAGCAACAACGGATCTACTCAGCCAACTGTTCGAAGCGCTCCAA
CAAGTCCGCTCAACTCTTCCGCTCAACATCATCATCTCCAACGGCTCAGAAACGAGGA
AAGGCGCTCCGTGAAGCGCGCGATCGAGCTCTTGCCGTTTCCGCTCGGGGAAGAACACTC
TGGAGCAGAGCGATCTTAGCTAATCGGATCAAACTGAAATTCGTAAACAGAGACGTCCT
CGAGCTACGATGGCGATTCGGCCATGACTACGGTGTTAGTAGCAGCAGCAACAGATCG
AGAAAACGGAGAGTGTGGGTGTTGAGATTGAATAAGAAGAGTATACCGGATGTTAACCGG
AAAGTACGTGTTCTAGGCCGGTTAGTTCCCGGTTGCGGTAAACAATCCGTACCGGTGATT
CTAGAAGAAGCAACTGATTATATTACAGGCTCTGGAGATGCAAGTGAGAGCCATGAACCTCT
TTAGTTCAGCTTCTCTCTCTACGGCTCAGCTCCTCCACCGATTGATGAGGTTAAAAT
CGTCTTTTAAATCTACCATCTCTCGATCTTTCACAGCTTATGTGTATATAGAAGATTCTG
GTTTGATTATAATCTGTAACCTACTCTTCCCAACCGCTGATTCTTCTCTGCTACAAGTAAA
AGTAAATTTGAACCGAGTCTTCCCATTTTACGATCCTCAAGTCTAAATTAAGTATATG
ATTGATTAATAAAGTCTTTACCATTAGGGTTC

>G2521 Amino Acid Sequence (domain in AA coordinates: 145-213)
MASLISDIEPPTSTTSDLVRRKKRSSASSAASSRSSASSVSGEIHARWRSEKQQRISAK
LFQALQQVRLNSSASTSSSPTAQKRGKAVREAADRALAVSARGRTLWSRAILANRIKLKF
RKQRRPRATMAIPAMTTVVSSSSNRSRKRRVSVLRLNKKSIIPDVNRKVRVLGRLVPGCGK
QSVPVILEBATDYIQALEMQVRAMNSLVQLLSSYGSAPPPI*

>G258 (60..983)
AGTGACCACCCTGCTGGTTAATCAACACCAAGAGACCTTGTAATATATAAGTTAGGAAGA
TGAGAGAGAAGTGGGAAATGAAAAGAGATGAAATGGGACATCGATGTTGTGGAAAACACA
AAGTGAAGAGAGGTCTTTGGTCTCCAGAGGAAGACGAGAAGCTTCTTCGTTATATCACCA
CTCATGGTCATCCTAGTTGGAGTTCGGTTCCAAAGCTTGCCGGGTTGCAGAGATGTGGGA
AGAGTTGCAGATTAAGGTGGATAAACTATCTAAGGCCTGATCTGAGGAGAGGTTTCGTTTA
ATGAGGAAGAAGAGCAGATTATCATCGACGTACATCGTATTCTTGGTAACAAATGGGCTC
AGATTGCTAAGCACTTACCTGGACGCACTGATAATGAAGTCAAGAACTTTGGAACTCAT
GCATTAAGAAGAAACTTCTTTCTCAAGGCTTAGATCCTTCTACACATAATCTTATGCCTT
CACACAAAAGATCTTCTTCTTCAAACAATAATAATCCCCAAGCCAAACAAAACGACGT
CCATCATGAAGAACCCTACTGATCTTGATCAATCAACCACTGCTTTTCAATCACAAACA
TCAATCCACCCACTTCCACTAAACCAAACTAAATCTCCTAACCAGACTACAATCC
CATCTCAAACCGTGATCCCTATCAATGATAACATGTCAAGTACTCAAACCATGATCCCTA
TCAATGATCCCATGTCAAGTCTTTTAGATGATGAGAATATGATTCTCTCACTGGTCAGATG
TTGATGGAATGGCGATCCACGAAGCTCCGATGTTGCCTAGTGATAAGGCAGTAGTGGGAG
TGGATGATGATGATCTCAACATGGACATTTTGTTTAACTACTCTTCTCTCTGCTTTTG
ATCTGATTTTGTCTCCATTTCTCCTCTGCAATGTCTATCGATTCTCAATCCCATGGATG
ATCTTGGCAGCTGGACCTTTAGCTTTTACTCTACAGC

>G258 Amino Acid Sequence (domain in AA coordinates: 24-124)
MREKWEKRDDEMGRHCCGKHVKRGLWSPEEDEKLLRYITTHGHPSSVVKLAGLQRCG
KSCRLRWINYLRPDLRRGSFNEEEEQIIIDVHRILGNKWAQIAKHLPGRTDNEVKNFWNS

CIKKKLLSQGLDPSTHNLMPSHKRSSSSNNNNIPKPNKTTSIMKNPTDLDQSTTAFSITN
INPPTSTKPNKLKSPNQTTIPSQTVIPINDNMSSTQTMIPINDPMSLLDDENMIPHWS
VDGMAIHEAPMLPSDKAVVGVDLNDLMDILFNTTPSSSAFDPDFASIFSSAMSIDFNPM
DLGSWTF*

>G280 (108..722)

AAGTTAATATGAGAATAATGAGAAAACCACTTTCCCAAATTGCTTTTTTAAATCCCTCCT
CACACAGATTCCCTTCCTTCATCACCTCACACACTCTCTACGCTTGACATGGCCTTCGATC
TCCACCATGGCTCAGCTTCAGATACGCATTATCAGAACTCCGTCGTTTTCTCTCCAC
CTTATCCTCAGATGATAATGGAAGCGATTGAGTCCTTGAACGATAAGAACGGCTGCAACA
AAACGACGATTGCTAAGCACATCGAGTCGACTCAACAACTCTACCGCCGTCACACATGA
CGCTGCTCAGCTACCATCTCAACCAGATGAAGAAAACCGTCAGCTAATCATGGTGAAGA
ACAATTATATGAAACCAGATCCAGATGCTCCTCCTAAGCGTGGTCGTGGCCGCTCCTCCGA
AGCAGAAGACTCAGGCCGAATCTGACGCCGCTGCTGCTGCTGTTGTTGCTGCCACCGTCG
TCTCTACAGATCCGCTTAGATCTCGTGGCCGCTCACCAGCCGAAAGATCCATCGGAGC
CTCCCCAGGAGAAGGTCATTACCGGATCTGGAAGGCCACGAGGACGACCACCGAAGAGAC
CGAGAACAGATTTCGGAGACGGTTGCTGCGCCGGAACCGGCAGCTCAGGCGACAGGTGAGC
GTAGGGGACGTGGGAGACCTCCGAAGGTGAAGCCGACGGTGGTTGCTCCGGTTGGGTGCT
GAATTAATCGGTACTTATGCAATTTCCGAATCTTTAGTTACTGAAAAATGGAATCTCTTA
GAGAGTAAGAGAGTGCTTTAATTTAGCTTAATTAGATTTATTTGGATTTCTTTTCACTATT
TGGATTGTAACTTTAGAATTTGTGTGTGTGTTGTTGCTTAGTCCTGAGATAAGATATAA
CATTAGCGACTGTGTATTATTATTACTGTCATTGTGTTATGTGAAACTTTGTTCTCTT
GTTGAAAAA

>G280 Amino Acid Sequence (domain in AA coordinates: 97-104,130-137-155-162,185-192)

MAFDLHHGSASDTHSSELPSFSLPPYPQMIMEAIESLNDKNGCNKTTIAKHIESTQQTLP
PSHMTLLSYHLNQMKKTGQLIMVKNYMKPDPDAPPKRGRGRPPKQKTQAESDAAAAAVV
AATVVDSTPPRSRGRPPKPKDPSEPPQEKVITGSGRPRGRPPKRPRTDSETVAPEPAAQ
ATGERRGRGRPPKVKPTVVAVPGC*

>G3 (16..477)

GTTTGTCTTTTATCAATGGAAGAGAACAAAGAGTCTACGATGAGAAAGAGAAGGCAG
CCACCTCAAGAAGAAGTGCCTAACCACGTGGCTACAAGGAAGCCGTACAGAGGGATACGG
AGGAGGAAGTGGGGCAAGTGGGTGGCTGAGATTCTGTGAGCCTAACAAACGCTACGGCTT
TGGCTTGGCTCTTACACAACCGATATCGCCCGCTAGAGCCTACGACGTGGCCGCTTTC
TACCTCCGTGGCCCTCCGCACGTCTCAACTTCCCTGATCTTCTCTTGCAAGAAGAGGAC
CATCTCTCAGCCGCCACCACCGCTGACATGCCCGCAGCTCTTATAAGGGAAAAAGCGCG
GAGGTCCGGCGCCAGAGTCGACGCTCTTCTAGCTTCTGCCGCTCCTTCGATGGCTCACTCC
ACTCCGCCGTAATAAAACCCGACTTGAATCAAATACCCGAATCCGGAGATATATAGTCA
ATTTATATACATGTAGTTTGTGTTTGTGATTAGAAAGATTACATTTACATACAAGATACA
CATAGATACTGGAAAAATAGGTATGTATACATTATATAAATTATCTTATGTATCAAAGAA
TTTTATAGATTCTGATTAGCTTTTTGTTTTGTTTTGATAAGAACTCTGATTAGTTGTC
CGGAGACAAAACCGCTAAGAGCAATCCATGAGAAGCTAGCGAGTGTTTTTTAGTTCAAG
TTGTAATATAAATGCATATTAATTCTTTAGTAATTTGT

>G3 Amino Acid Sequence (domain in AA coordinates: 28-95)

MEREQEESTMRKRRQPPQEEVPHVATRKPYRGIRRRKWKWVAEIREPNKRSRLWLGSY
TTDIAARAYDVAVFYLRGPSARLNFDDLQEDHLSAATTADMPAALIREKAAEVGAR
VDALLASAAPSMHSTPPVIKPDNLQIPESGDI*

>G343 (1..795)

ATGGACGTCTATGGETTATCTTACCAGACTTACTTGAATCGACGACCTTCTTGATTTC
TCCAACGAAGACATCTTCTCGCTTCTTCTCCGGTGGTTCCACCGCCGCTACTTCTCT
TCTTCTTTCCCTCTCTCTCAAAACCCTAGTTTCCACCACCACCATCTCCCTTCTCCGCC
GATCATCACTCCTTCTCCACGACATTTGCGTTCACAGTGATGACGACGCTCATCTTGAA
TGGCTTTTCGAATTCGTGGACGATTCTTTCGCTGATTTTCCGGCGAATCCATTAGGAGGA
ACTATGACTTCTGTCAAACTGAACTTCTTTCGGGGAAACCAAGAAGCAAACGATCA
AGAGCTCCTGCTCCTTTCGCCGGAACATGGTCTCCGATGCCACTGGAATCCGAGCATCAG
CAGCTTCACTCCGCCGCCAAATTCAAGCCAAAGAAAGAACAATCCGGCGGAGGAGGAGGA
GGAGGAGGAAGACATCAGTCATCGTCATCGGAGACTACGGAAGGAGGAGGAATGAGGAGA
TGTACTCACTGTGCATCGGAGAAAACGCCACAGTGGAGGACAGGACCACTTGGACCTAAA

ACACTATGTAACGCTTGTGGAGTCCGGTTTAAATCCGGTAGACTTGTACCGGAATATAGA
CCGGCTTCGAGTCCTACTTTTGTGTTTGAAGTCTCAAACTCTCACCAGAAAGTGATG
GAGCTTCGACGGCAGAAAGAAGTTATGAGACAACCACAACAAGTTCAACTTCATCACCAC
CACCACCCGTTTTAG

>G343 Amino Acid Sequence (domain in AA coordinates: 178-214)
MDVYGLSSPDLRLIDLLDFSNEDIFSASSSGGSTAATSSSSFPFPQNPFSHHHLPSSA
DHHSFLHDICVPSDDAAHLEWLSQFVDDSFADFPANPLGGTMTSVKTETSFPGKPRSKRS
RAPAPFAGTWSFMPLESEHQQLHSAKFKEQSGGGGGGGRRHQSSSEETTEGGGMRR
CTHCASEKTPQWRTPGLPKTLNACGVRFKSGRLVPEYRPASSPTFVLTHQHSNSHRKVM
ELRRQKEVMRQPQQVQLHHHHHPF*

>G363 (1..780)
ATGAGACCAATATTAGACCTCGAAATTGAAGCTTCATCGGGCAGTAGTAGCAGCCAAGTG
GCCTCAAACCTTGTCTCCGGTTGGGGAAGATTACAAACCAATCTCGCTGAATCTTAGCCTC
AGTTTCAACAACAACAACAATAATCTGGATCTTGAATCATCGTCTTTGACGCTGCCA
CTTTCGAGCAGAGTGAGAGTAGTAACCCGGAGCAGCAGCAGCAACAACAACCATCTGTA
TCAAAGAGAGTCTTCTCTGTAACTACTGCCAAAGGAAGTTCTATAGCTCTCAAGCGTA
GGTGGTCACCAAAACGCTCACAAACGTGAGAGAACACTCGCCAAACGCGCTATGCTATGG
GTCTTGTCTGGGTCTTCCCCGGTAGAGGATCAAGTAGCAATTATGCGGCTGCTGCCACAG
CAGCCGCTCTCGTGTGTCGCTTACCGAAGCGGAAACGGGAACATGACATCGTTCAGG
ACTTTGGGAATCCGGGCACATTCTCGGCGCAGCAGCTCAGCATGACAAGGCAGACACCA
GAAACACTTATTAGAAACATTGCCAGGTTCAACCAGGGGTATTTCGGTAATTGTATACCT
TTTACGTGGAGGACGACGAGGCCGAGATGCTCTGGCCGGGGAGTTTCCGGCAAGCTACG
AATGCGGTTGCGGTTGAAGCGGGTAATGATAATTTAGGTGAAAGAAAAATGGATTTCTTG
GACGTCAAGCAAGCGATGGATATGGAAGTTCTCTTCCAGATCTAACCTTGAAGCTTTGA
>G363 Amino Acid Sequence (domain in AA coordinates: 87-108)
MRPILDLEIEASSGSSSSQVASNLSPVGEDYKPISLNLSLNFNNNNNNLDLESSSLTLP
LSTSESSNPEQQQQQPSVSKRVFSCNYCQRKFYSSQALGGHQAHRERTLAKRAMLW
VLLGSSPVEDQVAIMRLLPQQLSCLPLHSGNGNMTSFRTLGIRAHSSAHDVSMTRQTP
ETLIRNIARFNQGYFGNCIPFYVEDDEAEMLWPGSFRQATNAVAVEAGNDNLGERKMDFL
DVKQAMDMESSLPDLTLKL*

>G370 (1..774)
ATGGACGAAACCAACGGACGAAGAGAAACTCACGATTTTCATGAACGTCAACGTTGAATCC
TTCTCTCAGCTTCTTTTCATCCGCGTACTCTCCCAAAGAAAAGCCGCCATTATTCGT
CTCTTCGGCCAAGAGCTCGTTCGGTGATAACTCCGACAACCTTATCCGCAGAACCTTCTGAT
CATCAAACCATACCAAGAACGATGAGAGCTCTGAGAATATCAAGGACAAAGACAAAGAA
AAAGATAAGGACAAAGACAAAGATAACAACAACAGGAGATTTCGAGTGCTACTACTGC
TTCAGAAACTTCCCAACTTCTCAAGCCCTAGGTGGACATCAAAACGCTCACAACGTTGAA
CGTCAACACGCCAAACGCGGTTCCATGACATCATACCTTCATCATCATCAGCCTCATGAC
CCTCACCACATCTACGGCTTCTCAACAACCACCACCACCGTCACTATCCGTCTTGGACG
ACGGAAGCTAGATCATACTACGGCGGAGGGGACATCAAACGCCGTCGTACTACTCAAGG
AATACTCTTGCTCTCTCTTCTTAACCCACCGACAATCAACGGAAGTCCTTTAGGTTTG
TGGCGGTGACCGCCTTCCACGTCAACAAATACTATTCAAGGCGTTTACTCATCTTCACCA
GCTTCAGCGTTTAGGTTCGATGAGCAAGAGACTAATAAGGAGCCTAATACTGGCCGTAC
AGATTGATGAAACCAATGTGCAAGATCATGTGAGTCTCGATCTTCATCTCTGA

>G370 Amino Acid Sequence (domain in aa coordinates: 97-117)
MDETNGRRRETHDFMNVNVEFSQLPFIRRTPPKEKAAIIRLFGQELVGDNSDNLSEPSD
HQTTTKNDESSENIKDKDEKDKDKDNNNNRRFECHYCFRNFPTSQALGGHQAHRERE
RQHAKRGSMTSYLHHHQPHDPHHIYGLNNHHHRHYPSTTEARSYYGGGHHQTPSYYSR
NTLAPPSSNPPTINGSPLGLWRVPPSTSTNTIQGVYSSSPASAFRSHEQETNKEPNWPY
RLMKPNVQDHVSLDLHL*

>G385 (37..2202)
TAGGGTTTGCTTTCAGTTTCCGGAGTATAAGAAAAGATGTTTCGAGCCAAATATGCTGCTT
GCGGCTATGAACAACGCAGACAGCAATAACCACAACCTACAACCAGGAAGACAACAATAAT
GAAGGATTTCTTCGGGACGATGAATTCGACAGTCCGAATACTAATCGGGAAGTGAGAAT
CAAGAAGGAGGATCAGGAAACGACCAAGATCCTCTTCATCCTAACAAGAAGAACGATAT
CATCGACACACCCAACCTTCAGATCCAGGAGATGGAAGCGTCTTCAAAGAGTGTCCTCAC
CCAGATGACAAGCAAAGGAAACAGCTAAGCCGTGAATTGAATTTGGAACCTCTTCAGGTC

>G385 Amino Acid Sequence (domain in AA coordinates: 60-123)
MFEPNMLLAAMNNADSNHHNHNHEDNNNEGFLRDDEFDSPNTKSGSENGEGSGNDQDPL
HPNKKKRYHRHTQLQIQEMEAFFKECPHPDDKQQRKQLSRELNLLEPLQVKFWFQNKRTQMK
NHHERHENSHLRAENEKLRNDNLRYREALANASCPNCGGPTAIGEMSFDEHQRLLENARL
REEIDRISAIAAKYVGKPVSNYPLMSPPLPPRPLELAMGNIGGEAYGNNPNDDLKSKITA
PTESDKPVIIDLVAAMEELMRMVQVDEPLWKSALDEEYARTFFRGIGRPRAGYRSEA
SRESADVIMNHVNIVEILMDVNQWSTIFAGMVSRAMTLAVLSTGVAGNYNGALQVMSAEF
QVPSPLVPTRETYFARYCKQQGDGSWAVVDISLDSLQPNPPARCRRRASGCLIQELPNGY
SKVTWVEHVEVDDRGVHNLKYHVMVSTGHAFGAKRWVAILDRQCERLASVMATNISSGEVG
VITNQEGRRSMLKLAERMVISFCAGVSASTAHTWTTLSTGTGAEDVRVMTRKSVDDPGRSP
GIVLSAATSFWIPVPPKRVDFDLRDENSRNEWDILSNGGVVQEMAHIANGRDTGNCVSL
RVNSANSSQSNNMILQESCIDPTASFVIYAPVDIVAMNIVLNGGDPDYVALLPSGFAILP
DGNANSAGPGDGGSLTLTFAFQILVDSVPTAKLSLGSVATVMNLIAC TVERIKASMSCET
A*

TATAAATCTTCGTTTC TACTTTTTTTCTTCCATAATATAGTCAATTCTGTTTCTTAATT
AGGGCTTCTTCTCTTGTGTTCTCCAATCTTTATTAGTTTATTTATTTATTTTGGTTATTG
TATACAAATGGCAATGGCTTTAAACATGAATGCTTACGTAGACGAGTTCATGGAAGCTCT
TGAACCATT CATGAAGTAACTTCATCTTCTTCTACTTCGAATTCATCAAATCCAAACC
ATTAAC TCCTAATTTCAATCCCTAATAATGACCAAGTCTTACCGGTATCTAACC AAACCGG
TCCGATTGGGCTAAACCAGCTCACTCCAACACAAATCCCTCCAAATT CAGACAGAGTTACA
TCTCCGGCAAAACCAATCTCGTCGTCGCGCTGGTAGTCATCTTCTCACC GCTAAACCAAC
CTCAATGAAGAAAATCGAGGTAGCAACTAAACCGGTTAAACTATACCGAGGCGTAAGACA
GAGGCAATGGGGTAAATGGGTAGCTGAGATTCTGGCTACCTAAAAACCGAACCCGGTTATG
GCTCGGTACGTTTCAAGACGGCTCAAGAAGCTGCATTAGTTCAGATCAAGCAGCTCATAA
GATCAGAGGAGACAACGCTCGTCTCAATTTCCAGACATTTGTTTCGTCAAGGACACTATAA

ACAGATATTGTCTCCGTCTATCAACGCAAAGATCGAATCCATCTGCAATAGTTCTGATCT
TCCACTGCCCTCAGATCGAGAAACAGAACAAAACAGAGGAGGTGCTCTCTGGTTTTTCCAA
ACCGGAGAAAGAACCAGAAATTTGGGGAGATATACGGATGCGGATACTCGGGCTCATCTCC
TGAGTCGGATATAACGTTGTTGGATTTCTCAAGCGACTGTGTGAAAGAAGATGAGAGTTT
CTTGATGGGTTTGCACAAGTATCCTTCTTTGGAGATTGATTGGGACGCTATAGAGAAACT
CTTCTGAATCCATTTTATCTTTTGGATTCAATTTGTCTCTAAATTGTAGAATTTTATTTTC
AGAGCTTTGTAAGGGAAGTTCTTGAATGAGAGTTGCAGAGGACTAGTGGAACTAACTCT
GTTTTCTTTTGTAAGTATTGTTTATAATGGGCCGTTGAATGGGCCTTATTGATTAAACA
GCCCAAGTTTTTAAAAAAAAAAAAAAAAAAAAAAAAA

>G439 Amino Acid Sequence (domain in AA coordinates: 110-177)
MAMALNMNAYVDEFMEALEPFMKVTSSTSNSSNPPLTPNFIPNNDQVLPVSNQTGPI
GLNQLTPTQILQIQTELHLRQNSRRRAGSHLLTAKPTSMKKIDVATKPKLYRGVRQRQ
WGKWVAEIRLPKNRTRLWLGTFTETAQEAALAYDQAAHKIRGDNARLNFPDIVRQGHYKQI
LSPSINAKIESICNSSDLPLPQIEKQNKTEEVLSGFSKPEKEPEFGEIYCGYSGSSPES
DITLLDFSSDCVKEDESFLMGLHKYPSLEIDWDAIEKLF*

>G440 (237..1301)
AAAAAATCACTGTTTCATAACACGTTTTTCTCTCTACCCACCAAAAAAATCTTTTGT
TCTTGTTTACCAAAAAATCTCGTGATAAATCTCTTCAAACTTTGTTTTATTTCTTCTGA
TTCTCTCGAAATCTCTCTCAACAAACCCAGAACTTTCTTGTATTGCAAGCTTTTCTTC
CTTTTATATTCTTCATTTTGTATGCGAATATAGAGAGAGTCCATAAAGAAACAGTAATGG
ACGAATATATTGATTTCGACCATTGAAGTACACAGAGCACAAGACTTCAATGACTAAAT
ACACCAAAAAGTCATCGGAAAACTTTCCGGTGGTAAGTCATTGAAAAAGGTTAGTATTT
GTTATACTGATCCTGACGCAACAGATTCAATGACGAAGACGAAGAAGATTTCTTGT
TTCTCTCGCCGGAGAGTCAAAAGATTCTGTTAACGAGATCACTGTTGAGCCTAGCTGTAACA
ACGTCGTCACCGGAGTTTCGATGAAAGATAGAAAGAGACTCTCTTCTCTCCGATGAAA
CTCAATCTCCGGCTCGAGTCGTCACGTCCTAATAACAAAGTTTCAGTCTCCGGTCAGA
TAAAGAAAGTTCCGTTGGTGTAGACAACGCCATGGGGGAAATGGGCGGCGGAGATTAGAG
ATCCGGAGCAACGTCGGAGGATTTGGCTCGGGACTTTTGAGACGGCGGAGGAAGCTGCCG
TGGTTTATGATAACGCCGCTATAAGACTCCGTGGACCGGACGCTTTAACTAATTTCTCCA
TACCGCCTCAAGAAGAGGAAGAAGAAGAAGAACCGGAACCGGTTATTGAGGAGAAACCGG
TTATTATGACGACGCCAACCAACAACATCGAGTTCTGAATCAACTGAAGAAGATTTAC
AACATCTCTCATCTCTTCTCGGTTCTCAATCACCGGTCAGAAGAGATTCAACAAGTAC
ACAACCGTTTAAATCAGCTAAACCCGAACCGGGGGTTTCAAATGCACCATGGTGGCATA
CCGGGTTTAAATACCGGTTTAGGTGAATCAGACGATTCAATTTCTTTGGATACTCCGTTTC
TTGACAACTATTTTCAATGAATCACCACCAGAGATGTCAATATTTGACCAACCAATGGATC
AAATTTTCTGTGAAAATGATGATATCTTCAATGATATGTTGTTCTTGGGTGGTGAAACTA
TGAACATTGAAGATGAGTTAACAAAGTTCTAGTATCAAAGATATGGGTTCAACGTTTAGTG
ATTTTGTATGATTCAATGATATCAGATCTATTAGTTGCTTAATATGATGATGAGAGTGAAG
AAGAAACCATCAAGCAATATCTATGGTGTGACTGAAAAATTTTGGTGTACTTTTTTTT
CTTTCATAAGTTCATGAGCTTTTGTGTTCTTTTTTTTAAATAATTTATTAGTTTGTCA
GGAGCTTGTAACACAGTTTGGAGAAATAGTGGAATAATAGTTTAAATTAATAAAAAAAAAA
AAAAAA

>G440 Amino Acid Sequence (domain in AA coordinates: 122-189)
MDEYIDFRPLKYTEHKTSMTKYTKSSEKLSGGKSLKKVSICYTDPDATDSSSEDEEDF
LFPRRRVKRFVNEITVEPSCNNVTVGSMKDRKRLSSSDETQSPASSRQRPNNKVSVSG
QIKKFRGVRQRPWGKWAEEIRDPEQRRRIWLGTFETAEEAAVVYDNAAIRLRGPDALTNF
SIPPQEEEEEEPEPVIEKPVIMTPTPTSSSESTEEDLQHLSSPTSVLNHRSEEIQQ
VQQPFKSAKPEPGVSNAPWWHTGFNTGLGESDDSFPLDTPFLDNYFNESPPEMSIFDQPM
DQIFCENDDIFNDMLFLGGETMNIIEDELTSSSIKDMGSTFSDFFDLSLIDLLVA*

>G5 (417..1421)
TTTTTTTTTGCATCTCCCCCTAATCTGTTGTTTCTCGCTTCTTCTTCTGTTAATCATC
TGCTTTTCAAAAAGAAAGAAAAAAGAAAAATTCGATTTCTGGGTTTGTGTTTGTGCATACA
GAAAAAATCAAGCTTATGAATTTGTGTTAATTTTGTGTTTAAATTTGAAAGGCAGGTT
TTTTTCAGAACGAGATCGTTTTTCAAATTTCTTCTGATTTTACCTCTTTTTTCTTCTTA
GATTTTAGTGAATCGAGGGTGAAATTTTGTATCCCTCTTTTCGGATCTACACAGAGGTT
GCTTATTTCAAACCTTTTAGATCCATTTTCTGATTTTCTCGGAAAAATCCCTGTTTC
TTTACTTTTTTATAAGTCTCAGGTTCAATTTTTCGGATTCAAATTTTATTTTAAATGG

CAGCTGCTATGAATTTGTACACTTGTAGCAGATCGTTTCAAGACTCTGGTGGTGAAGTCA
TGGACGCGCTTGTACCTTTTATCAAAAGCGTTTCCGATTCTCCTTCTTCTTCTTGCAG
CGTCTGCGTCTGCGTTTCTTACCCCTCTGCGTTTCTCTCCCTCCTCTCCCGGTTATT
ACCCGGATTCAACGTTCTTGACCAACCGTTTTCATACGGGTCGGATCTTCAACAAACCG
GGTCATTAATCGGACTCAACAACCTCTCTTCTCTCAGATCCACCAGATCCAGTCTCAGA
TCCATCATCTCTTCTCCTCCGACGCATCACAACAACAACCTCTTCTCGAATCTTCTCA
GCCAAAGCCGTTACTGATGAAGCAATCTGGAGTCGCTGGATCTTGTTCGCTTACGGTT
CAGGTGTTCTTCTGAAGCCGACGAAGCTTTACAGAGGTGTGAGGCAACGTCAGTGGGAA
AATGGGTGGCTGAGATCCGTTTGGCGAGAAATCGGACTCGTCTCTGGCTTGGGACTTTTG
ACACGGCGGAGGAAGCTGCGTTGGCCTATGATAAGGCGGCGTACAAGCTGCGCGGCGATT
TCGCCCCGGCTTAACCTTCCCTAACCTACGTCATAACGGATTTACATCGGAGGCGATTTCG
GTGAATATAAACCTCTTCACTCCTCAGTCGACGCTAAGCTTGAAGCTATTTGTAAGCA
TGGCGGAGACTCAGAAACAGGACAAATCGACGAAATCATCGAAGAAACGTGAGAAGAAGG
TTTCGTCGCCAGATCTATCGGAGAAAGTGAAGGCGGAGGAGAATTCGGTTTCGATCGGTG
GATCTCCACCGGTGACGGAGTTTGAAGAGTCCACCGCTGGATCTTCGCGGTTGTCGGACT
TGACGTTTCGCTGACCCGGAGGAGCCGCGCAGTGAACGAGACGTTCTCGTTGGAGAAGT
ATCCGTCGTACGAGATCGATTGGGATTTCGATTCTAGCTTAGGGGCAAAATAGGAAATTC
GCCGCTTGCAATGGAGTTTGTGAAATTGCATGACTGGCCCAAGAGTAATTAATTAAT
ATGGATTAGTGTAAATTTTCGTATGTTAATATTTGTATTATGGTTTGTATTAGTCTCTCT
GTGTCGGTCCAGCTTTCGGGTTTTCGTCAGGCTCGACCATGCCACAGTTTTTCATTTTATG
TAATCTTTTTTCTTTTGTCTTATGTAATTTGTAGCTTCAGTTTCTTCATCTATAATGCA
ATTTTATTATGATTATGTG

>G5 Amino Acid Sequence (domain in AA coordinates: 149-216)
MAAAMNLYTCSRFSQDSGGLMDALVPFIKSVSDSPSSSSAASASAFSLHPSAFSLPPLPG
YYPDSTFLTQPSYGSDDLQQTGSLIGLNNLSSSQIHQIQSQIHHPPLPPTHNNNNNSFNL
LSPKPLLMLKQSGVAGSCFAYGSGVPSKPTKLYRGVRQRHWGKWVAEIRLPRNRTRLWLGT
FDTAEEAALAYDKAAYKLRGDFARLNFNLRHNGFHIGGDFGEYKPLHSSVDKLEAICK
SMAETQKQDKSTKSSKKREKKVSSPDLEKVKAEENSVSIGGSPPVTEFEESTAGSSPLS
DLTFADPEBPQWNFTFSLEKYPSEYIDWDSILA*

>G550 (1..1374)

ATGGCTGATCCGGCGATTAAGCTCTTTGGAAAGACGATTCCTTTACCTGAGCTTGGTGT
GTTGATTCTTCTTCTAGCTATACCGGATTTTAAACGAAACTCAGATTCTGTTCGGTTA
TCAGATTTCGTGTACCGGCGATGATGATGATGAAGAGATGGGTGATTCCGGTTTAGGACGA
GAAGAAGTGATGATGTTGGTGATGGTGGAGGAGAGAGCGAGACTGATAAAAAGGAAGAA
AAAGATAGTGAGTGTGAGGAAGAGTCATTGAGGAATGAATCTAATGATGTTACTACTACT
ACATCGGTATAACTGAAAAACGGAACAACAAAAGCTGCAAGACGAATGAAGAGTCA
GGTGGTACTGCTTGGCTCTCAAGAGGGGAAGTTAAAGAAACCTGATAAGATTCTACCGTGT
CCGCGATGTAACAGCATGGAAACCAAGTTCTGTTACTACAACAACATAATGTTAACC
CCTCGCCATTTCTGCAAGAAATGTCAGAGATATTGGACAGCTGGTGGAACGATGAGGAAT
GTTCCGGTTGGTGTGAGGAGACGTAAGAATAAGAGTCCAGCTTCTCATTATAACCGTCAT
GTAAGTATAACATCTGCGGAAGCTATGCAGAAGGTGGCGAGAAGTATCTTCAACATCCT
AATGGTGCAATCTTCTCACTTTTGGCTCTGATTCTGTGCTTTGTGAATCTATGGCTTCT
GGATTGAATCTTGTGAGAAGTCATTGTTGAAGACACAACTGTATTGCAAGAACCCAAT
GAAGGCTTGAAGATTACGGTTCCGTTAAACCAGACAAACGAAGAAGCTGGAACAGTCAGC
CCGTTACCAAAAAGTTCCATGCTTTCCAGGACCACCACCAACTTGGCCTTACGCTTGGAAC
GGAGTTTTCGTGGACGATTTTACCGTTTACCTCCACCGGCTTACTGGAGCTGCCCGGGG
GTTTCACCGGGGGCATGGAACAGCTTACATGGATGCCACAACCAATTACCATCTGGT
TCCAATCCAAATCTCCTACACTAGGTAAACATTACAGTGACGAGAACGCTGCTGAACCA
GGAACCGCTTTTGTGAAACCGAGTCACTTGGTAGGGAGAAAAGCAAAACCGAGAGATGC
TTGTGGGTTCCCAAGACGCTGAGGATTGATGATCCAGAGGAAGCTGCTAAAAGTTCCATC
TGGGAAACATTAGGGATCAAAAAGACGAAATGCGGATACTTTCGGAGCTTTCAGATCA
TCAACCAAGAAAAAGCAGTCTTCTGAAGGAAGACTTCCGGGAAGAAGACCGGAGTTG
CAAGCGAATCCTGCTGCTCTTCTAGGTCAGCAAACTTCCATGAGAGCTCATAG

>G550 Amino Acid Sequence (domain in AA coordinates: 134-180)
MADPAIKLPGKTIPLPELGVVDSSSYTGFLTETQIPVRLSDSCTGDDDDDEEMGDSGLGR
EEGDDVGDGGGESETDKKEEKDSECQESLRNESNDVTTTSGITEKTETTKAAKTNEES
GGTACSQEGKLKPKDILPCPRCNSMETKFCYNNYNNVNPQRHFCCKCQRYWTAGGTMRN

VPVGAGRRKNKSPASHYNRHVSITSAEAMQKVARTDLQHPNGANLLTFGSDSVLCESMAS
GLNLVEKSLKKTQTVLQEPNEGLKITVPLNQTNEEAGTVSPLPKVPCFPGPPTWPHYAWN
GVSWTILPFYPPPAYWSCPGVSPGAWNSFTWMPQPNPSGSPNSPTLGKHSRDENAAEP
GTAFDETESLGREKSKPERCLWVPKTLRIDDPPEAAKSSIWETLGIKKIDENADTFGAFRS
STKEKSSLSSEGRPLGRRPELQANPAALSRSANFHES*

>G670 (28..1152)

CACAGCATTGCAGCTGTGAATAACTAAATGGGGAGACATTCTTGCTGTTACAAACAAAAG
CTGAGGAAAGGGCTTTGGTCTCCTGAAGAAGACGAGAAGCTTCTTACTCACATCACCAT
CACGGCCATGGCTGCTGGAGCTCTGTCCCTAAACTCGCTGGTTTGCAGAGATGTGGGAAG
AGTTGTCGACTCGAGCAGATCTGGTACCGCCGACTAAGATGGATCAATTACTTGAGACCT
GATTTAAAGAGAGGAGCTTTTTCTCCTGAAGAAGAGAATCTCATCGTCGAACCTTCATGCC
GTCCTTGGAACAGATGGTCACAGATTGCGTCAAGGCTTCCGGGTAGAACCGACAACGAG
ATCAAGAATCTATGGAACCAAGCATCAAGAAGAACTGAAACAAAGAGGCATTGACCCA
AACACACACAAGCCCATCTCTGAAGTGGAGAGTTTTAGCGACAAAGACAAACCAACA
AGCAACAACAAAGAAGCGGTACGATCACAAGTCTCCTAGTTCTCTTCTGCGACTAAC
CAAGACTTCTTCTCGAAAGGCCATCTGATTTATCCGACTACTTCGGATTTTCAAGGCTT
AACTTCAACTCCAATCTAGGACTCTCTGTTACAACCTGATTCTTCACTCTGCTCGATGATT
CCGCCGAGTTTAGCCCCGGGAACATGGTTGGTTCTGTCTTCAGACACCAGTATGCGTA
AAGCCCTCGATTAGTCTTCTCCCGACAACAACAGTTCGAGTCTATCTCCGGAGGAGAT
CATGTGAAATTGGCTGCACCAAACTGGGAATTTTCAACAAACAATAATACCTCAAAT
TTCTTCGACAATGGCGGATTCTCATGGTCTATCCCAAATCTTCTACTTCTTCTTCAAA
GTCAAACCAATCATAACTTCGAAGAAATAAAATGGTCAGAGTATTTGAACACACCGTTC
TTCATAGGGAGTACTGTACAGAGTCAAACCTCTCAACCAATCTACATCAAATCAGAAACA
GATTACTTAGCCAATGTTTCAAACATGACAGATCCTTGAGCCAAAACGAGAACTTGGGC
'ACAACCTGAACTAGTGACGTGTCTTCCAAGGATCTTCAGAGAATGGCCGTCTCTTTTGGT
CAGTCCCTTTAGCTTTTTCTTTCTTTCTTTCTTATTCTAACAGATGTAGAGAACATAA
AGATATACAAATACATACAATGTCAATACGTACAGTGGATTAAAGTGTCTGTATATTTT
ATGGGCGAGCTGTCTTTATTTTTATGTTTAAAAA

>G670 Amino Acid Sequence (domain in AA coordinates: 14-122)

MGRHSCCYKQKLRLKGLWSPEEDEKLLTHITNHGHGCWSSVPKLAGLQRCGKSCRLEQIWY
RRLRWINYLRPDLKRGAFSPBEENLIVELHAVLGNRWSQIASRLPGRDNEIKNLWNSSI
KKLKLQKRGIDPNTHKPISEVESFSDKDKPTTSNNKRSNDHKSPPSSSATNQDFFLERPS
DLSDYFGFQKLNFSNLGLSVTTDSSLCSMIPPQFSPGNMVGSVLQTPVCVKPSISLPPD
NNSSSPISSGDHVKLAAPNWEFQTNNNNTSNFFDNGGFSWSIPNSSTSSSQVKPNHNFEE
IKWSEYLNTPFFIGSTVQSQTSQPIYIKSETDYLANVSNMTDPWSQENENLGTETSDVFS
KDLQRMVVSFGQSL*

>G760 (175..1878)

TGCTTAATTCCAATGCCATCGTGATCGATTCACTCTCTCTCTCTTCCAATTTTCCCA
ATTCTTTTAAACCCCTAATTTTTCAGATATCTGATTATCTCTTGTATTCTTCTACTC
GATTGCTCCCATAAAAACCTTACTTTCTTCAAGTTCTGGTTTTCACCGATTGATGGGT
CGTGGCTCAGTGACGTGCTGCTCCTGGGTTCGGTTTTACCCGACGGATGAGGAACCT
GTTGCTACTACCTTAAGCGTAAGGTCTGCAACAAACCTTTAAGTTCGATGCTATTTCC
GTCACCGACATATACAAGTCTGAGCCTTGGGATCTACCAGATAAGTCGAAGCTGAAAGT
AGAGACTTGGAATGGTACTTCTTTAGTATGCTGGATAAGAAGTACAGTAATGGTTCCAAG
ACGAATCGTGCTACGGAGAAAGGGTATTGGAAGACGACTGGGAAAGATCGGGAGATTCTGT
AATGGTTCAAGAGTCGTTGGGATGAAGAAGACACTTGTTTATCACAAGGGTCGAGCTCCT
CGTGGTGAAAGGACCAATTGGGTTATGCATGAGTATCGGCTTTCTGATGAGGACTTGAAG
AAAGCTGGTGTGCCACAAGAAGCATATGTGTTATGTAGGATATTCAGAAAAGTGGTACG
GGTCTTAAGAATGGGAGCAGTATGGTGCTCCTTATCTTGAGGAGGAGTGGGAAGAAGAT
GGAATGACTTATGTACCTGCTCAAGATGCTTTCAGTGAAGGATTGGCTTTGAATGATGAT
GTTTATGTGATATTGATGACATTGACGAGAAGCCCGAAATCTGGTGGTCTATGATGCC
GTTCTTATTCTACCTAACTATTGTATGGGGAATCAAGTAACAATGTTGAATCAGGCAAT
TACTCAGACTCTGGAAATTACATTCAACCAGGAAACAATGTTGTCGACTCTGGTGGGTAC
TTTGAACAACCAATTGAACTTTTGGAGGAAGATCGGAAGCCTATTATACGGGAGGGTAGC
ATTGAGCCTTGTCTCTGTTTCCAGAGGAACAAATTGGCTGTGGTGTGCAAGACGAAAT
GTGGTGAATCTGGAATCTTCCAACAATAATGTGTTTGTAGCTGATACATGCTACAGTGAC
ATTCTTATTGATCATAACTATTTACCCGATGAGCCATTCTGGATCCTAATAACAATCTT

CCACTCAACGATGGTCTGTACCTGGAAACGAATGATCTCAGCTGTGCTCAACAAGATGAT
TTTAACCTCGAAGATTATCTCAGCTTCTTTGATGATGAGGGTTTGACTTTTGACGATTCT
CTATTAATGGGACCTGAAGATTTTCTTCCCAACCAAGAAGCCCTTGACCAGAAACCTGCC
CCTAAAGAATTGGAGAAGGAGGTCGCAGGAGGCAAGAGGCAGTGGAGGAAAAGGAAAGT
GGCGAAGGATCTTCTTCAAAAACGAGCCACATGCTTGGAGCCATTCTACTCCATCTTCATT
TACCCATTTCTCAAAAAGACGAGCCACATGCTTGGAGCCATTCTACTCCATCTTCATT
GCTTACAGTTCCAAACAAAGGACGCAATGCGTCTACACGCAGCACAATCTTCTGGTTCA
GTTACAGTGAAGTGCAGGTATGATGAGAATATCAAACATGACTCTAGCAGCGGACAGCGGT
ATGGGCTGGTCATATGACAAGAACGGTAACCTCAACGTAGTCCCTTCATTTCGGGGTAGTC
CAACAGGATGATGCGATGACTGCCTCGGGAAGCAAGACAGGAATTACGGCGACAAGAGCT
ATGTTAGTCTTCATGTGTTTATGGGTTCTCTACTCTCTGTTAGCTTCAAAATAGTAACC
ATGGTGTCTGCTCGGTAATAGGATCAAAGTTGAATCGTCTCAAAGACTTTTTTTGGTGT
TGTACCTCTCCAATCATATAGCCTTTAAGTTTGGCAGTGCTTGGCTGCTCAATATTTAAA
TTTTAAAAA

>G760 Amino Acid Sequence (domain in AA coordinates: 12-156)
MGRGSVTS LAPGRFHPHTDEELVRYLKRKVCNKPFDKDAISVTDIYKSEPWDLPDKSKL
KSRDLEWYFFSMLDKKYSNGSKNTRATEKGYWKTGKDREIRNGSRVVGMMKTLVYHKGR
APRGERTNWMHEYRLSDEDLKKAGVPQEAYVLCRIFQKSGTGPKNGEQYGAPYLEEWE
EDGMTYVPAQDAFSEGLALNDDVYVDIDIDEKPENLVVYDAVPILPNYCHGESSNNVES
GNYSDSGNYIQPGNNVVDGGYFEQPIETFEEDRKPIIREGSIQPCSLFPEEQIGCGVQD
ENVVNLESSNNNVFVADTCYSIDIPIDHNYLPDEPFMDPNNNLPLNDGLYLETNDLSAQQ
DDFNFEYDLSFFDDEGLTFDDSLMGPEDFLPNQEAALDQKPAPKELEKEVAGGKEAVEEK
ESGEGSSSKQDTFKDFDSAPKY PFLKKTSHMLGAIPTPSSFASQFQTKDAMRLHAAQSS
GSVHVVTAGMMRISNMTLAADSGMGWSYDKNGNLNVVLSFGVVQQDDAMTASGSKTGITAT
RAMLVFMCLWVLLLSVSFKIVTMVSAR*

>G831 (92..1987)

TTCTTTTCATCGTTGTGTCTATTATAAATATATGTCAATTTGGTTTCTAAAAAATCTACC
ATTGATTGATTGATTTTTTTTTCTTTAAGAGATGAATTTATTTACAAGAATCTCATCTCG
GACTAAGAAGGCCAATCTTTACTACGTAACCTAGTTGCTCTTCTCTGCATCGCTAGCTA
CCTTCTCGGTATTTGGCAAAACACGGCGGTTAATCCACGCGCCGCTTCGATGATTGAGA
CGGTACACCGTGCGAGGGATTACACGACCTAATTCTACGAAAGATCTCGACTTCGACGC
GCATCACAACATTCAAGATCCACCTCCGGTGACGGAACCGCCGTTAGTTTCCCGTCGTG
TGCCGCCGCGTTGAGCGAGCACACGCCATGCGAAGACGCGAAGCGATCGTTGAAATCTC
GAGGAGAGATTGGAGTATAGGCAAAGGCATTGTCCCGAGAGAGAAGAAATCTTGAAGTG
CAGAATTCGGCGCCGTACGGTTACAAAACGCGGTTCCGATGGCCGCGAGTCGTGACGT
GGCGTGGTTTCGCTAATGTGCCTCACACGGAGCTTACGGTTGAGAAAAAGAAATCAGAATTG
GGTCCGGTACGAGAATGATCGGTTTTGGTTCCCTGGTGGAGGTACGATGTTCCACGTGG
CGCTGATGCTTACATTGATGATATCGGACGGTTGATTGATCTCAGCGACGGCTCTATCCG
TACAGCCATCGATACCGGTTGCGGGGTGGCTAGCTTCGGTGCTATCTTTTATCAAGAAA
CATTACAACGATGTCTGTCACCAAGAGACACACGAAGCTCAAGTCCAGTTCGCACT
CGAGCGTGGTGTGCGCGGATGATCGGAATCATGGCTACAATCCGCTACCGTACCCCTC
TAGAGCCTTTGATTTAGCACATTGCTCTCGTTGCCTTATTCCGTGGGGCCAAAACGATGG
GGCTTACTTGATGGAGGTGGATAGGGTTTTAAGACCAGGAGGGTACTGGATACTTTCTGG
ACCGCCGATTAATTGGCAGAAACGGTGGAAAGGGTGGGAACGGACCATGGATGATTTGAA
TGCAGAGCAGACTCAGATCGAGCAGGTCGCGAGAAGCTTGTGTTGGAAGAAAGTTGTTCA
AAGAGATGATCTTGCTATTTGGCAAAAACCCCTTTAACCACATTGACTGTAAGAAAACCG
AGAGGTTTTGAAAAATCCGGAGTTTTGTGCTCATGATCAAGATCCCGACATGGCCTGGTA
TACGAAGATGGATTGTGTTTGACACCATTAAGTTGATGACGCTGAGGATCTAAA
GACGGTGGCCCGGAGGGAAGGTAGAAAAGTGGCCGGCTAGATTAAACGCGATTCTCCGAG
AGTAAACAAAGGCGCTCTCGAGGAAATCACACCTGAAGCTTCTTGGAGAACACGAACT
GTGGAACAGAGAGTTTCTTATTACAAGAAGTTAGATTACCAGTTGGGTGAAACCGGGAG
ATACAGAAACTTAGTCGACATGAACGCTTACCTCGGTGGATTTCGCGCGGCTCTAGCGGA
TGATCCGGTCTGGGTCAATGAACGTTGTCCCGGTGAGGCTAAGCTCAATACGCTCGGTGT
CATCTACGAGCGTGGTCTAATCGGAACGTATCAAACTGGTGTGAAGCCATGTGACGTA
TCCAAGAACGTATGATTTTATCCATGCTGACTCGGTTTTACATTTGTACCAAGGTCAATG
TGAACCGGAGGAGATATTGTTGGAGATGGACCGAATTTCTAGACCGGGTGGTGGTGTGAT
TATAAGAGATGACGTGGACGTTTGGATCAAGGTTAAGGAATTAACCAAAGGATTAGAATG

GGAAGGTAGAATTGCTGACCACGAGAAGGGTCCTCATGAAAGAGAGAAGATTTACTATGC
GGTGAAACAGTATTGGACCGTTCTCTGCGCCTGATGAAGATAAAACAACACTAGTGCTCT
CTCCTGATTTTTGAGTTTTTTTTTTTCTTACAATGTTTTTTTTTTTTTTTCAATTTTT
TATACAACAATAAATTCTCAATAATTGTTGTGTCGCGGCCG

>G831 Amino Acid Sequence (domain in AA coordinates: 470-591)
MNLFRISRRTKKANLYVTVLVALLCIASYLLGIWQNTAVNPRAAFDDSDGTPCEGFTRP
NSTKDLDFDAHNIQDPPPVTEAVSFSPSCAAALSEHTPCEDAKRSLKFSRERLEYRQRH
CPEREELKCRIPAPYGYKTPFRWPASRDVAWFANVPHELTVKKNQNWVRYENDRFWF
PGGGTMFPRGADAYIDDIGRLIDLSDGSIRTAIDTGGCVASFGAYLLSRNITMSFAPRD
THEAQVQFALERGVAMPIMIGIMATIRLPYPSRAFDLAHCSRCLIPWGQNDGAYLMEVDRVL
RPGGYWILSGPPINWQKRWKGWERTMDDLNAEQTQIEQVARSLCWKKVVQRDDLAIWQKP
FNHIDCKKTREVLKNPEFCRHDQDPMAWYTKMDSCLTPLPEVDDAEDLKTAVAGGKVEKW
PARLNAIPPRVKNKGALEBITPEAFLENTKLWKQRVSYKKLDYQLGETGRYRNLDVMNAY
LGGFAAALADDPVWVMNVVPEAKLNTLGVIERGLIGTYQNWCEAMSTYPRTYDFIHAD
SVFTLYQGQCEPEEILLEMDRILRPGGGVIIRDVDVLIKVELTKGLEWEGRIADHEKG
PHEREKIYYAVKQYWTVPAPDEDKNNTSALS*

>G864 (503..1534)
TGCAAAAACATTTTCTGTCTCTCTCTGCCCCAAATTTTTTTTCTTTCCAGGAATATTTTC
CTAGAAAAACCCAAGCAAAGCTTTAACCCTCTCTCTCCAAAGTAGCATCTTCCTCTT
TTTCTATTTCTCTTCTCTCTCTTATCTCTCTCTCGTTTGTGAACGATTCTTAAGAAT
ATAACCAAAGCCCTTTTCTCTTCTTCAACTTTCCGGGAAAAATCTTCACGCAGCAAG
GTTTCTCTCTCGGCTCTCCAGTGTTTTTCCGGCCTTTTGTCTTTCTATAAAAAA
TTCGCGTCTCTTAAGAAACTTTTTCCACCTAGAGAAGAAGAAGATCACTCTTGTTG
TTCAAGTTTCTCTTTAATAAAAAATCCATCTTTATTCTTTGTCTCTTTCTCTTTTGC
TTTCCCTAATCTCTATGTTATAACACACAGAGAGAAACAAAGTCACAGTCTCGAGTCAA
AAACAGAGAATACGAAAGAAAAATGGAAGCGGAGAGAAGAAATGGTTCTACCGAGAATCAA
ATTACAGAGACACAAACCAACACGACAACAATCGTATCGGAGTTAACCAACACTCACCA
AACCAGGATTCTTCGTATCTCAGTCACTGACCCAGACGCTACTGATTCTCCAGTGACGA
CGAAGAAGAAGACATCAACGCTTTGTCTCTAAACGCCGTCGTGTTAAGAAGTTTGTCAA
CGAAGTCTATCTCGATTCCGGTGCTGTTGTTACTGGTAGTTGTGGTCAAATGGAGTGCAG
GAAGAGACAAAAGAGAGCGGTTAAATCGGAGTCTACTGTTTCTCCGGTTGTTTCAGCGAC
GACGACTACGACGGGAGAGAAGAAGTTCCGAGGAGTGAGACAGCGTCCATGGGGAAAAATG
GGCGGCGGAGATAAGAGATCCGTTGAAACGTGTACGGCTCTGGTTAGGTACTTACAACAC
GGCGGAAGAAGCTGCTATGGTTTACGATAACGCCGCTATTTCAGCTTCGTGGTCCCCGACGC
TCTGACTAATTTCTCAGTCACTCCGACAACAGCGACGGAGAAGAAAGCCCCACCACCGTC
TCCGGTGAAGAAGAAGAAGAAAAACAACAAAGCAAAAAATCCGTTACTGCTTCTTC
CTCCATCAGCAGAAGCAGCAGCAACGATTGTCTCTGCTCTCCGGTGTCTGTTCTCCGATC
TCCTTTCGCCGTCGACGAATTCTCCGCATTCTTCATCACCAGTCGCGGCCGTTGTAGT
CAAGGAAGAGCCATCCATGACAACGGTATCTGAAACTTTCTCTGATTTCTCGCGCCCTT
GTTCTCAGATGATGACGTGTTTCGATTTCCGGAGCTCAGTGGTTCCCGACTATCTCGGCGG
CGATTTATTTGGGGAAGATCTATTCACGGCGGATATGTGTACGGATATGAACCTCGGATT
CGATTTCCGATCCGATTATCCAGCTGGCACATGGAGGACCATTTTCAAGATATCGGGGA
TCTATTCCGGTCCGATCCTCTTTTAGCTGTTTAATAATATTTTAAATAAATAAATAGTTA
TACCGGCCGTTACTAAACGGAACCGGAGAAAGTTTTGTATACCGGTGACATAAAATCTCG
GTTATGTTTCGTAATCTTTTTTCTTTGTTATATATAAAATATGAATGAACTGAATTA
TGTAAGTTAATGGTGATAATTATTAACGTTTAAAGTTTTGAAAAAAAAAAAAAAAAAAAA
AAAAAA

>G864 Amino Acid Sequence (domain in AA coordinates: 119-186)
MEAEKMKVLPRIKFTEHKNTTTIVSELNTHQTRILRISVTDPDATDSSSDDEEEHQR
FVSKRRRVKFKFVNEVYLDGSAVVTGSCGQMESKKRQKRAVKSESTVSPVVSATTTTTGEK
KFRGVRQRPWGKWAELIRDPLKRVRLWLGTYNABEAAAMVYDNAIQLRGPDALTNFSVT
PTTATEKKAPPPSPVKKKKKKNNKSKSVTASSSISRSSNDCLSPVSVLRSPFAVDEF
SGISSSPVAVVVKEEPSMTTVSETFSDFSAPLFSDDDVDFRSSVVPDYLGDLFGEDL
FTADMCTDMNFGDFGSGLSWHMEDHFQDIGDLFGSDPLLAV*

>G884 (31..1575)
TTTTTTTTTGTTTGTAAATTTTGGGGATCGATGTCGGAAGGAAGAGCTCCGTCGACA
TCGAAGTCCACCGGAGCTCCGTCGCGTCCGACTTTATCTCTCTCCACGGCCGTTAGT

GAGATGTTCTTTAACGGTGGCGTTGGATTCACTCCTGGTCCGATGACTCTGGTCTCTAAT
ATGTTCCCTGATTCCGATGAGTTTAGGTCTTTCTCTCAGCTTCTCGCTGGAGCCATGTCT
TCTCCAGCGACTGCAGCTGCTGCTGCTGCTGCTGCGACGGCTAGTGATTACCAGAGACTT
GGTGAAGGGACTAATAGCTCTAGTGGTGATGTTGACCCGAGATTCAAGCAAAACAGACCA
ACCGGTTTGATGATTTCTCAATCTCAATCGCCGTCGATGTTACCGTACCGCCTGGTTTA
AGTCCAGCTATGTTGCTCGATTACCAAGCTTTTTGGGTCTTTCTCTCCCGTTCAGGGA
TCATATGGAATGACACATCAGCAAGCTCTAGCTCAAGTCACTGCTCAAGCAGTTCAAGCC
AATGCCAATATGCAACCAACAGAGTACCCTCCTCCCTCTCAAGTTCAATCATTTTCA
TCGGGTCAAGCGCAGATCCCGACCTCGGCTCCACTACCAGCTCAAAGAGAAACCTCAGAT
GTAACCATCATAGAGCACAGGTCACAACAGCCTCTAAATGTTGACAAACAGCTGATGAT
GGCTATAACTGGCGAAAATATGGGCAAAAGCAAGTTAAAGGTAGCGAGTTTCCACGAAGC
TATTACAAGTGTACTAATCCAGGATGTCTGTCAAGAAGAAGGTTGAGAGATCTCTTGAT
GGACAAGTAACGGAGATTATCTACAAAGGTCAGCACAAATCATGAACCTCCTCAAAACACT
AAGCGAGGTAACAAAGATAACACCGCGAATATAAATGGGAGTTCGATAAATAACAATCGC
GGGAGTTCTGAATTGGGGGCATCAGTTTCAAACCTAATAGCTCCAACAAGACTAAGAGA
GAGCAACATGAAGCAGTAAGTCAAGCTACGACAACAGAGCACTTGTCTGAGGCAAGTGAC
GGTGAAGAAGTTGGTAATGGAGAAACTGATGTGAGAGAGAAAGATGAGAATGAGCCTGAT
CCCAAGAGAAGTACAGAAGTTCCGATTTCAGAACAGCTCCTGCTGCTTCACATAGA
ACTGTGACAGAGCCTAGAATTATTGTCCAAACGACGAGTGAAGTTGATCTTCTAGATGAT
GGATATAGGTGGCGTAAATATGGACAGAAAGTTGTCAAAGGGAATCCTTATCCGAGGAGC
TACTACAAGTGCACAACACCAGGATGTGGTGTGAGGAAACATGTAGAGAGAGCAGCAACA
GATCCAAAAGCTGTAGTAACAACATATGAAGGAAAACATAACCATGACCTTCCCGCTGCT
AAATCAAGCAGCCATGCCGCTGCAGCGGCACAGTTAAGGCCAGATAATCGACCTGGCGGT
TTGGCTAACTTAAATCAACAGCAGCAGCAACAGCCCGTTGCGCGGCTAAGGCTTAAAGAA
GAGCAAACTTGAAGAGAAGAACTCTTGACCGTTTTTCATTACAAAAGCTTTCAAAT
TCCACTCACACACTTGTCTGAAAATCTAGCAGTTTGCAGGAAAGAAACAGCTTCAAGAG
GTTGTAGTTCTCTATGTTCTGGTGTAAACCTTAAAGCTTTTTAGGGTTTTTCAGATTTT
TGTTTACTAATACTGTATGTGAATCTTTTGTACATGAGGAAGAAAATTACAGGGGGATA
TTTTGTGTTGTATCTTTTGTGTTATGTTTTCAGTAAAAGATAGGTCTTACATTTTGTGTA
AAAAAAAAAAAAAAAAAAAA

>G884 Amino Acid Sequence (conserved domain in AA coordinates:227-285, 407-465)

MSEKEEAPSTSKSTGAPSRPTLSLPPRPFSEMFNGGVGFSPGPMTLVSNMFPDSDEFRS
FSQLLAGAMSSPATAAAAAAATASDYQRLGEGTNSSSGDVPFRKQNRPTGLMISQSQS
PSMFTVPPGLSPAMLLDSPSFLGLFSPVQGSYGMTHQQLAQVTAQAVQANANMQPQTEY
PPPSQVQSFSSGQAQIPTSAPLPAQRETSDVTTIEHRSQQPLNVDPKADDGYNWRKYGQK
QVKGSEFPRSYKCTNPGCPVKKKVERSLDGQVTEIIYKQHNHEPPQNTKRNKJONTAN
INGSSINNNRGSSELGASQFQTNSSNKTREQHEAVSQATTEHLSEASDGEEVNGNETD
VREKDENEPDKRRSTEVRISEPAPAASHRTVTEPRIIVQTTSEVDLLDDGYRWRKYGQK
VVKGNPYPRSYKCTTPGCGVRKHVERAATDPKAVVTTYEGKHNHDLPAKSSSHAAAAA
QLRPDNRPGGLANLNQQQQPVARLRLKEEQTT*

>G898 (161..772)

GAAAAAAGATTCAAAAACCTAGATTTCAAAAAATCGATTGGCTGTCAAATTTCTCTCC
GGCGATTTTCTCGAGTGAAATTCGGCTCAAGGTGATTATAGCGATCATCGAATCAAAT
GATTGAAGAGGTACAAAGGTTAGTTACTTTGAGCTGAAAGATGAACACGTGAGAGGTGAG
AGTACCTCGAGGAAATCGACGAGGAAAGCTGTGATTGATCTGAATGCGGTACCTGTTGA
TCAAGAAGGGACCTCTGCTTCTGTTAGAATCTTACGGTGCCTATTACACCGTCTCAGCC
TGCTCCTACGATGATTGATGTCGATGCTATTGAGGATGATGTTATTGAATCATCCGCTAG
TGCTTTTGTGTAAGCTAAAAGCAAATCAAGAAATGCACGTCGGAGACCTTTGATGGTTGA
TGTAGAGTCAGGAGGTACGACTAGATTCCCTGCCAACATAAGCAACAAACGAGAAGGAT
TCCTTCTAGTGAATCTGTCTGACTGTGAGCATGCCTCTGTAAATGATGAAGTCAACAT
GTCTTCGAGAGTGTCTAGATCAAAGGCTCCAGCTCCTCCACCAGAAGAGCCAAAGTTTAC
ATGTCCAATCTGCATGTGTCCCTTTACGGAGGAGATGTCAACCAAGTGCGGTCAATCTT
CTGCAAGGGATGTATAAAGATGGCAATATCTCGCCAGGGCAAATGCCCTACTTGTAGGAA
AAAGGTTACTGCAAAAGAGCTGATTCCGAGTTTTCTTCCAACCACTAGATGAGTGGTCCG
GCAACATCACCAGCCACCCTGTCTAATGGTTTATCAGACTATCCTCCTATTCACTTTGGA
ACATTGAAGGGACTTCGTTGACTTGGTATTTTTGAATATTTGCTTTGTTGGAAGAGAAA
TATTCAAGTATCAAGAAGCCAGAAGGCCCTATCATTCGATGGATATCATTGGTAATAACT

CTTTGTTTTTAGTTGTTGTTCTATGTAATTTAGGTCTCTGCAAACCTCTCAGTCGATACT
CTTCTCTCTTGATAGATGATAAGATATATGGAATAATATTGAATCTTTACTA
AAA

>G898 Amino Acid Sequence (domain in AA coordinates: 148-185)
MNTSEVRVPRGNRRRKAVIDLNAVVDQEGTSASVRLTVPITPSQPAPTMIDVDAIEDD
VIESSASAFEAESKSRNARRRPLMVDVESGGTTRFPANISNKRRRIPSSSEVIDCEHAS
VNDEVNMSRVSRSKAPAPPPEEPKFTCPICMCPFTEEMSTKCGHIFCKGCIKMAISRQG
KCPTCRKKVTAKELIRVFLPTTR*

>G900 (1..648)
ATGGGGAAGAAGAAGTGCAGTTATGTTGTGGTGTAGCGAGAATGTATTGTGAGTCAGAT
CAAGCGAGTTTATGTTGGGATTGTGACGGTAAAGTTCACGGAGCTAATTTCTGGTGGCG
AAACACATGCGTTGTCTTCTATGTAGCGCGTGTCTAGTCACACACGCCCTTGGAAAGCTTCT
GGGCTGAATCTTGGCCCAACTGTTTCTATCTGTGAGTCTTGTTTAGCTCGTAAGAAGAAT
AACAACAGCTCCCTCGCCGGGAGGATCAGAATCTTAACCAAGAAGAAGAGATCATTGGT
TGTAACGACGGAGCTGAGTCTTATGATGAGGAAAGCGATGAGGATGAAGAAGAAGAAGAA
GTGGAGAATCAGGTTGTTCCGGCTGCGGTGGAGCAAGAACTTCCGGTGGTGAGTTCGTCTG
TCTTCGGTTAGTAGTGGTGAAGGAGATCAGGTGGTGAAGAGACGAGACTTGATTTGGAT
CTTAACCTCTCCGATGAGGAGAACCAATCTAGACCATTGAAAAGATTATCGAGAGACGAA
GGTTTGTCAAGATCAACTGTTGTGATGAATAGCTCAATCGTGAAATTACACGGAGGGAGG
AGAAAAGCAGAGGGATGTGATACATCATCGTCTCTTCTGTTTTATTGA

>G900 Amino Acid Sequence (domain in AA coordinates: 6-28, 48-74)
MGKKKCELCGGVARMYCESDQASLCWDCDGKLVHGANFLVAKHMRCLLCSACQSHTPWKAS
GLNLGPTVSICESCLARKKNMNSLAGRDQNLNQEEIIGCNDGAESYDEESEDDEEEEEE
VENQVVPAAVEQELPVSSSSSVSSGEGDQVVKRRLDLNLSDEENQSRPLKRLSRDE
GLSRSTVVMNSSIVKLHGGRRKAECDTSSSSSPY*

>G913 (108..806)
CATTCAAAAACATCATATATATACACAAACACACTTTGATACAACAAAAAACAGAAC
ACAAACAAAAACACATTGTAACATTAGTTTAAGCATTAAGCTTCTTTATGTGCAATAATA
ATAATTCTCCGACCACCGTGAATCAAGAAACGACGACGTCTCGTGAAGTCTCAATCACAT
TGCCTACTGATCAATCTCCTCAAACTCACCAGGATCATCTTCTTCTCTTCCACCGAGAC
CTTCCGGTGGATCACCAGCGAGAAGAACGGCGACTGGATTATCCGGCAAGCACTCTATTT
TCAGGGGGATTCCGACTACGTAACGGAAATGGGTATCGGAGATTAGAGAGCCACGTAAAA
CGACAAGAATTTGGCTCGGACTTATCCGGTACCGGAGATGGCTGCCGCCGCTTACGACG
TGGCTGCGTTAGCTTTAAAAGGACCCGACGCCGTTTTGAATTTTCTGTTTAGCTTTGA
CTTACGTGGCTCCGTTTCAAACCTCTGCTGCGGATATAAGAGCGGCTGCTAGTAGAGCAG
CGGAGATGAAGCAACCGGATCAGGGTGGGGATGAGAAGGTATTGGAACCGGTTCAACCCG
GCAAAGAGGAAGAATTAGAAGAAGTGTCGTGTAACCTCGTGTTCGTTGGAGTTTATGGATG
AGGAAGCGATGTTGAATATGCCGACTTGTGTGACGGAGATGGCTGAAGGGATGTTGATGA
GTCCACCGAGAATGATGATACATCCGACGATGGAAGATGATTCCGCCGAGAATCATGAAG
GAGATAATCTTTGGAGTTATAAATGAATCCATTGAAGCTGCTCTCTTTTTTATTGTTTTC
CGGTGGAATGAGATTTTCCCCCTTTTTTTTTTTCTTTTTGGGTGCGTGTT

>G913 Amino Acid Sequence (domain in AA coordinates: 62-128)
MSNNNSPTTVNQETTSREVSITLPTDQSPQTSPPSSSPSPRPSGGSPARRTATGLSG
KHSIFRGIRLRNGKWVSEIREPRKTRIWLGTYVPPEMAAAAYDVAALALKGPDAVLNFP
GLALTYVAPVSNSAADIRAAASRAEMKQPDQGGDEKVLPEVPQPGKEEELEEVSCNSCSL
EFMDEEAMLNMPDLLTEMAEGLMSPPRMMIHPTMEDDSPENHEGDNLSYK*

>G937 (45..1046)
TGAAAAAGTTTGAATTTTTAATTGGAATCGAGAAAAATAAAAAATGGGTTCTTTAGGTG
ATGAGCTTAGTTTGGGATCGATCTTTGGGAGAGGAGTTTCGATGAATGTTGTGGCGGTTG
AGAAAGTTGATGAACATGTTAAGAAGCTTGAAGAAGAGAAGAGAAAGCTCGAAAGTTGTC
AACTTGAGCTTCTCTGTCTTTGCAGATTTTAAACGATGCGATTTGTATCTGAAGGATA
AGAGATGTTTCAGAGATGGAGACTCAACCATTGTTGAAAGATTTCATTTCTGTTAATAAAC
CTATTCAAGGAGAAAGAGGAATAGAATTGCTGAAAAGAGAGGAGCTAATGAGGGAGAAGA
AGTTTCAGCAATGGAAGCTAATGATGATCACACTAGTAAGATCAAGAGCAAGCTTGAGA
TTAAGAGAAATGAGGAGAAATCTCTATGTTGTTGATTCCAAAGGTGGAACTGGTTTAG
GCCTCGGTTTAAAGTTTCGAGTTCCGATAAGAAGAAAGGGATTGTTGCCTCATGTGGCTTTA
CTTCTAACTCTATGCCACAACCAACACCAGCAGTACCACAACAACCAGCATTCTTTA

AGCAGCAAGCTTTACGGAAGCAAAGAAGGTGTTGGAATCCAGAGTTGCATCGCCGATTTG
TCGATGCATTGCAACAGCTAGGTGGACCGGGAGTGGCAACTCCTAAACAAATTAGAGAAC
ATATGCAAGAAGAAGGCTTAACCAATGATGAAGTCAAGAGTCATTTACAGAAATACAGGT
TACACATCAGGAAGCCAAATTCGAATGCGGAGAAACAATCAGCAGTTGTTTATAGGGTTTA
ACTTGTGGAATCTTTCAGCACAGATGAAGAAGAGACATGTGAAGGAGGAGAATCATTGA
AGAGAAGCAATGCGCAATCAGATTCTCCTCAAGGTCCTTTGCAGTTACCGTCTACAACAA
CAACAACCTGGTGGAGATAGTAGCATGGAAGATGTTGAAGATGCTAAGTCTGAGAGCTTTC
AACTGGAGAGATTGAGATCACCATAAATCTCAAGAAACCAACTCTTGATCACGGTTTTG
TTATTTTGGATTCTACTATATCTATTAGTAGTGAATGAGAACAATAATTATAGAAAGG
TTTATAGATATATATATAGAGAAAAAGAGAGAGTGGGATGGTTCAAATTATTTGCAGA
>G937 Amino Acid Sequence (conserved domain in AA coordinates:197-246)
MGSGLDELSTLGSIFGRGVSMNVVAVEKVDHVKKLEEEKRKLESCQLELPLSLQILNDAL
LYLKDKRCSEMETQPLLKDFISVKNKPIQGERGIELLKREELMREKKFQQWKANDDHTSKI
KSKLEIKRNEEKSPMLLIPKVETGLGLGLSSSIRRKIVASCFTSNMPQPPTPAVPQ
QPAFLKQOALRKQRRCWNPFLHRRFVDALQQLGGPGVATPKQIREHMQEEGLTNDEVKSH
LQKYRLHIRKPNNSNAEKQSAVVLGFNLWNSSAQDEEETCEGGESLKRSNAQSDSPQGPLQ
LPSTTTTTGGDSSMEDVEDAKSESFQLERLRSP*

>G960 (63..1538)

TACCGTCGACCCACGCGTCCGAGTGATTCAAAGTCGGAAAGAAACCCTAAAGAAGAGGA
TTATGGGTGCTGTATCGATGGAGTGCCTTCTTTAGGTTTCAGATTCAGACCTACCGATG
AAGAGCTCGTCAATCACTACCTCCGTCTCAAGATCAACGGACGTCCTCCGATGTCCTG
TCATCCCTGATATCGATGTCTGCAAATGGGAACCTTGGGATCTTCTGCTCTCTCGGTGA
TTAAGACGGATGATCCAGAGTGGTCTTTTTCTGCCCTCGTGATCGGAAATACCTAATG
GTCATCGCTCTAACAGAGCAACTGACTCTGGCTATTGGAAGCTACTGGTAAAGATCGTA
GCATCAAGTCTAAGAAGACTTTAATCGGTATGAAGAAGACTCTTGCTCTTATCGTGGAC
GAGCTCCTAAAGGTGAGCGGACTAATTGGATTATGCACGAGTATCGTCCCACTCTTAAGG
ATCTTGATGGCACTTCCCCTGGCCAAAGCCCTTACGTTCTTTGTGCGCTCTTCCACAAGC
CTGATGATCGGGTTAATGGTGTCAAGTCCGATGAAGCAGCTTTTACGGCCAGCAACAAAT
ACTCACCTGATGATACATCATCTGATCTTGTTCAAGAAACACCTTCTCTGATGCTGCTG
TTGAGAAACCATCAGATTATTCAGGTGGATGCGGTTATGCTCATAGTAATAGTACCGCAG
ATGGGACAATGATTGAGGCACCTGAAGAGAATCTTTGGTTATCTTGTGACCTTGAAGATC
AAAAGGCACCACTACCGTGTATGGATTCTATATATGCTGGTGATTTCAGTTACGATGAGA
TTGGATTCCAATTCAAGATGGTACCAGCGAACCAGATGTATCACTAACAGAATTGTTGG
AGGAGGTGTTCAATAACCCTGATGACTTCTCTTGCGAGGAATCGATCAGTCGAGAGAATC
CAGCAGTCTCACCAATGGGATATTTTCATCTGCTAAAATGCTGCAGTCTGCAGCACCAG
AGGATGCTTTCTTCAACGACTTCATGGCTTTCACTGATACAGATGCTGAGATGGCGCAAT
TGCAGTATGGTTTCAGAAAGGTGGAGCTTCTGGTTGGCCAAGTGACACTAATTCATACTATA
GTGATTTGGTTTCAGCAAGAGCAAATGATCAATCATAACACAGAGAACAACCTCACAGAAG
GGAGAGGGATAAAGATCCGGGCTCGACAGCCTCAGAACCGGCAGAGTACAGGATTGATAA
ACCAGGGTATTGCTCCAAGGAGAATCCGTCTGCAGCTGCAGTCTAACTCTGAAGTAAAAG
AACGAGAGGAGGTGAATGAAGGACACACTGTTATTCCCGAGGCCAAAGAAGCTGCAGCTA
AATACTCAGAGAAGAGTGGTTCTTTGGTTAAACCTCAAATAAAGCTCAGGGCGCGGGGAA
CTATAGGCCAAGTAAAGGAGAGAGATTTGCAGACGACGAGGTACAGGTGCAGAGCACAA
AGAGAGAGAGAGAGAGAATCAAATGTAGTTTAAATGTAATTAGGGATGATGCAATGTTAGC
ATGTTTGTGTGTTGTAACCTTAAAACTTATTTAGGAATCTGATAAAAGTTACTGTTGAAA
AAAGAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

>G960 Amino Acid Sequence (domain in AA coordinates: 13-156)

MGAVSMESLPLGFRFRPTDEBLVNHYLRLLKINGRHSVDVRVIPDIDVCKWEPWDLPALSVI
KTDDPEWFFFCPRDRKYPNGHRNRRATDSGYWKATGKDRSISKKTTLIGMKKTLVFYRGR
APKERTINWIMHEYRPTLKDLDGTSFGQSPYVLCRLFHKPDDRVRNGVKSDEAFTASNKY
SPDDTSSDLVQETPSSDAAVEKPSDYSGGCGYAHNSNADGTMIEAPEENLWLSCDLEDQ
KAPLPCMDSIYAGDFSDEIGFQFQDGTSEPDVSLTELEEVFNPNPDDFSCEESISREN
AVSPNGIFSSAKMLQSAAPEDAFFNDFMAFTDTDAEMAQLQYGESEGGASGWPSDTNSYYS
DLVQQEQMINHNTENNLTTEGRGIKIRARQPQNRQSTGLINQGIAPRRIRLQLQSNSEVKE
REEVNEGHTVIPEAKEAAAKYSEKSGSLVKPQIKLRARGTIGQVKGERFADDEVQVQSTK
RERERIKCSLM*

>G991 (6..533)

GAAAAATGGAAGAAGAAAAGAGATTGGAGCTAAGGCTAGCTCCTTGTACCAATTCA
 CTTCCAACAACAACATCAATGGATCTAAACAAAAAGCTCGACCAAGAAACATCATTCC
 TTTCCAATAACAGGGTTGAGGTAGCTCCAGTGGTGGGATGGCCCGGTGAGATCATCCC
 GGAGAAACCTAACGGCACAACCTAAAGGAGGAGATGAAGAAGAAGGAGAGTGATGAAGAGA
 AGGAATTGTACGTTAAGATCAACATGGAAGGAGTTCCAATAGGAAGAAAAGTCAACCTTT
 CAGCTTATAACAACCTACCAACAGCTTTCACATGCCGTTGACCAACTCTTCTCTAAGAAAAG
 ATTCGTGGGATCTAAACAGACAATACACTTTGGTCTACGAAGACACTGAAGGAGATAAAG
 TTCTGGTCGGGGATGTTCTTGGGAGATGTTTGTATCTACTGTAAAGAGGTTGCATGTTT
 TAAAGACCTCCACGCCTTCTCACTCTACCTAGAAAACATGGCAAGGAATAGAGAGAGG
 TTGGCCAAAACATCAGTTCGATGGTTTGTTTTAAATGTAATTTTGTGGAACTAATGG
 GGTTTGGCTTTTGATTTACTGGTTTTCTTTTCACTTATGTACTAGGTTTTTGTCTGCTAT
 GTTATTTCTTGTTTTGGTTGTAAATATGCTGTTCTGTTTAAAGAAATCGGGGGTTAGTATGT
 TATCGTGTGTATAAAAAATAGTGTAAAGCACGTAAGTTGATTACAAAAAATAAAAAA
 AAAAAA

>G991 Amino Acid Sequence (domain in AA coordinates: 7-14,48-59,82-115,128-164)

MEEKRLELRLAPPCHQFTSNNNINGSKQKSSTKETSFLSNRVEVAPVVGWPPVRSSRR
 NLTAQLKEEMKKKESDEEKELYVKINMEGVP IGRKVNLSAYNNYQQLSHAVDQLFSKKDS
 WDLNRQYTLVYEDTEGDKVLVGDVPWEMFVSTVKRLHVLKTSFAFSLSPRKHGKE*

>G748 (98..1444)

CCACGCGTCCGCACCTCTCCCAATCTCTTCTTTAACAACAAAAAATACACAGAGA
 CATAGAGAGAAGAAGACGGAACAGAGGCTCCAAAAAATGATGATGGAGACTAGAGATCC
 AGCTATTAAGCTTTTCGGTATGAAAATCCCTTTTCCGTCGGTTTTTGAATCGGCAGTTAC
 GGTGGAGGATGACGAAGAAGATGACTGGAGCGGCGGAGATGACAAATCACCAGAGAAGGT
 AACTCCAGAGTTATCAGATAAGAACAACAACACTGTAACGACAACAGTTTTTAACAATTC
 GAAACCCGAAACCTTGGACAAAGAGGAAGCGACATCAACTGATCAGATAGAGAGTAGTGA
 CACGCTGAGGATAATCAGCAGACGACCTGATGGTAAAACCTAAAGAAACCGACTAA
 GATTCTACCGTGTCCGAGATGCAAAAGCATGGAGACCAAGTTCTGTTATTACAACAATA
 CAACATAAACACGCTCGTCATTCTGCAAGGCTTGTGAGAGATATGGAGTCTGGAGG
 GACTATGAGGAATGTTCTGTGGGGGCGAGACGTCGTAAGAACAAGCTCATCTTCTCA
 TTACCGTCACATCACTATTTCCGAGGCTCTTGAGGCTGCGAGGCTTGACCCGGGCTTACA
 GGCAACACAAGGGTCTTGAGTTTTTGGTCTCGAAGCTCAGCAGCAGCACGTTGCTGCTCC
 CATGACACCTGTTATGAAGCTACAAGAAGATCAAAAGGTCTCAAACGGTGCTAGGAACAG
 GTTTCACGGGTTAGCGGATCAACGGCTTGTAGCTCGGGTAGAGAAATGGAGATGATTGCTC
 AAGCGGATCCTCTGTGACCACCTCTAACAATCACTCAGTGGATGAATCAAGAGCACAAAG
 CGGCAGTGTGTTGAAGCACAAATGAACAACAACAACAATAACATGAATGGTTATGC
 TTGATCTCCAGGTGTTCCATGGCCTTACACGTGGAATCCAGCGATGCCCTCCACAGGTTT
 TTACCCGCTCCAGGTATCCAATGCCGTTTACCCTTACTGGACCATCCCAATGCTACC
 ACCGCATCAATCCTCATCGCTATAAGCCAAAAGTGTTCAAATACAACTCTCCGACTCT
 CGGAAAGCATCCGAGAGATGAAGGATCATCGAAAAGGACAATGAGACAGAGCGAAAACA
 GAAGGCCGGGTGCGTTCTGCTCCGAAAACGTTGAGAATAGATGATCCTAACGAAGCAGC
 AAAGAGCTCGATATGGACAACATTGGGAATCAAGAACGAGGCGATGTGCAAAGCCGGTGG
 TATGTTCAAAGGGTTTGATCATAAGACAAAGATGTATAACAACGACAAAGCTGAGAACTC
 CCCTGTTCTTTCTGCTAACCTGCTGCTCTATCAAGATCACACAATTTCCATGAACAGAT
 TTAGAGTTACATATGTATATATATATGATGATTGATGTATGATAGATGATACTGG
 AGAATGATGAGTTTTTTGAGAATCAAACCTTTTCTTTCTAGTGATTGCCTTTATTCC
 TTTACATGTTTTTGGTCTCTGTACACTATTTGATTACCTTTTTTACTTTCTTTCTTCAT
 TTGTCAGGAAATGTTGGAAGATAACATTAATGGTAAAAAGTTGGTGTGGACCGTTGTTGC
 GTTGGCATTTCAAAAAATAAAAAA

>G748 Amino Acid Sequence (domain in AA coordinates: 112-140)

MMETRDPAIKLFGMKIPFPVVFESAVTVEDEDDWSGGDDKSPEKVTPELSDKNNNNC
 NDNSFNNSKPELTLDKEEATSTDQIESDTPEDNQQTTPDGKTLKKPTKILPCPRCKSMET
 KFCYNNYNINQPRHFCKACQRYWTAGGTMNRNVPVGAGRRKNKSSSSHYRHITISEALEA
 ARLDPLQLQANTRVLSFNGLEAQQHVAAAPMPVMKLQEDQKVSNGARNRFGHGLADQRLVAR
 VENGDDCSSGSSVTTSSNHSVDESRAQSGSVVEAQMNNNNNNNMNGYACIPGVWPYTW
 PAMPPPGFYPPPGYPMPFYPTWTIPMLPPHQSSSPISQKCSNTNSPTLGKHPRDEGSSKK
 DNETERKQKAGCVLPKTLRIDDPNEAAKSSIWTTLGIKNEAMCKAGGMFKGFDHKTMY
 NNDKAENSPVLSANPAALSRSHNFHEQI*

>G247 (1..660)

ATGAGAATGACAAGAGATGGAAAAGAACATGAATACAAGAAAGGTTTATGGACAGTGGAA
GAAGACAAGATCCTCATGGATTATGTCCGAACATCATGGCCAGGGCCACTGGAACCGCATC
GCCAAGAAAACCTGGGCTCAAGAGATGTGGGAAAAGCTGTAGGTTGAGATGGATGAACCTAC
TTAAGCCCTAATGTTAACAGAGGCAATTTTACTGACCAAGAAGAAGATCTCATCATCAGA
CTCCACAAGCTCCTCGGCAACAGATGGTCGTTGATAGCGAAAAGAGTTCCGGGAAGAACA
GACAACCAAGTAAAGAATTACTGGAACACACATCTCAGCAAGAACTTGGTCTCGGAGAT
CATTCAACTGCCGTCAAAGCCGCATGCGGTGTAGAGTCTCCACCGTCTATGGCCCTTATA
ACCACAACGTCCTCCTCTCATCAAGAGATCTCCGGTGGAAAAAATTCAACTCTAAGGTTT
GACACTTTAGTTGACGAATCCAACTCAAACCAAATCCAACTAGTCCACGCAACACCA
ACTGACGTAGAAGTTGCAGCTACGGTTCCAAATCTGTTGATACCTTTTGGGTCTTTGAA
GACGACTTCGAGCTTAGTTCACTCACTATGATGGATTTTACTAATGGGTATTGCCTTTGA
>G247 Amino Acid Sequence (domain in AA coordinates: 15-116)
MRMTRDGKEHEYKGLWTV EEDKILMDYVRTHGQGHWNRIAKKTGLKRCGKSCRLRWMNY
LSPNVNRGNFTDQEEDLIIRLHKLLGNRWSLIAKRVPGRTDNQVKNYWNTHLSKKLGLGD
HSTAVKAACGVESPSPMALITTTSSSHQEISGGKNSTLRFDTLVDESKLKP KSKLVHATP
TDVEVAATVPNLFDTFWVLEDDFELSSLTMMDF TNGYCL*

>G585 (111..2039)

CTCTCAAACATTTCTCTGTTTGTTCGGCGAAAACGGCAACTGTTTCATCAAATGACAAA
CACAAAAACCTTAACATCTAGTTTGTATCCTCTCTGATACCTCAAAAAAATGGATGAAG
AAACAATGGCTACCGGACAAAACAGAACAACTGTGCCAGAGAATCTGAAGAAACACCTCG
CAGTTTCAGTTTCGAAACATTCATGGAGTTATGGTATCTTTTGGTCTGTCTCTGCTTCTC
AGTCTGGAGTTT TAGAATGGGGAGATGGATACTATAATGGAGATATCAAAAACGAGGAAGA
CGATTCAAGCTTCGGAGATCAAAGCTGATCAGCTTGGTCTACGGAGGAGCGAGCAGCTTA
GCGAGCTTTACGAGTCTCTCTCCGTGCTGAATCTTCTTTCAGGCGTGTCTGCCGGAT
CTCAAGTCACCAAGAGAGCTTCCGCCGCCGACCTTTCACCGGAAGATCTCGCCGACACCG
AGTGGTACTATTTGGTTTGTATGTCTTTCGTCTTCAACATTGGTGAAGGAATGCCTGGAC
GGACGTTTGC AAAACGGTGAACCGATATGGTTGTGCAACGCTCATACGGCGGATAGTAAAG
TGTTTAGCCGTCTCTTCTAGCAAAAAGTGCTGCGGTTAAGACAGTGGTTTGCTTCCCGT
TCCTTGGAGGAGTCGTTGAGATTGGTACCACAGAACATATTACGGAAGACATGAATGTAA
TACAATGCGTGAAGACATCATTCCTCGAAGCCCCGTGATCCGTACGCTACAATATTACCAG
CAAGATCCGATTATCACATCGACAACGTTCTTGATCCGCAACAGATTCTAGGCGACGAGA
TTTACGCGCCTATGTTCAAGTACGAGAGCCTTTTCCAACAGCTTCTCCGAGCAGAACTACCA
ACGGTTTCGATCAAGAACATGAACAAGTAGCAGATGATCATGATTCTTTTCATGACCGAAA
GAATCACTGGAGGAGCTTCTCAGGTGCAAAAGCTGGCAGCTCATGGACGACGAGCTTAGTA
ACTGCGTTTACCAGTTCGCTAAATTCCAGCGATTGCGTCTCTCAAACGTTTGTGTAAGGGG
CGGCTGGACGGGTTGCTTACGGTGCAAGAAAGAGTAGAGTTCAAAGACTAGGGCAAATTC
AAGAGCAACAGAGAAATGTGAAGACATTGTCAATTTGATCCAAGAAACGACGACGTTCAAT
ACCAAAGTGTGATCTCAACGATTTTTAAGACCAACCATCAGTTAATTCTCGGACCGCAGT
TTCGAAACTGCGATAAACAGTCAAGCTTCACTAGGTGGAAGAAATCATCGTCATCATCAT
CAGGAACCGCCACGGTCACGGCACCATCACAGGAATGTTAAAGAAAATTATTTTCGATG
TTCCGCGAGTGCACCGAAGAGAGTTAATGTTGGACTACCAGAAGCCAGAGATGAAA
CTGGGAACCATGCGGTTT TAGAGAAGAAGCGCCGAGAAATTGAACGAACGGTTTCATGA
CCTTGAGAAAAATCATTCCGTCAATCAACAAGATCGATAAAGTATCGATTCTTGACGATA
CGATAGAGTATCTTCAAGAACTCGAGAGACGGGTTCAAGAACTAGAATCTTGACAGAGAAT
CAACCGATACAGAGACTCGTGGGACGATGACGATGAAGAGGAAGAAACCATGCGACGCAG
GAGAAAGAACATCAGCTAATTGCGCAAAATAATGAACAGGAAATGGGAAGAAGGTGTCGG
TTAACAATGTTGGTGAAGCCGAGCCAGCAGATACCGGTTTACTGGTTTAAACCGATAATT
TAAGGATCGGTTTCGTTTGGTAATGAGGTGGTTATTGAGCTTAGATGTGCTTGAGAGAAG
GAGTATTGCTTGAGATAATGGATGTGATTAGTGATCTCCATTGGATTCTCATTCGGTTC
AATCCTCGACCCGAGACGGTTTGCTCTGCTTAACCGTCAATTGCAAGCACAAAGGGTCAA
AAATAGCGACACCAGGAATGATCAAGAAGCACTTCAAAGGGTTCATGGATCTGTTGAA
GACTACTTAGTTAAAATTGACAGCAAAGAAAAAACATTCCCGGTTTGGTTTCTATTCTTT
GGTTTTCTTCTAACCGGGTTTTAGGAATTAATGTTATGTTTATCATTGTTTTTGTGTTT
TTTTTGTGTCCTTTTTTCCGTTTGCTTAACGTAGGTGAAGAGGAACATACACTATGCGTA
TTTTGTTTGAGGTAGATTATTTAAGGTATTAGTAATAGTAATAGCCAGTTTAGATGAT
TTTGTGTTCTTTTGTGTT

>G585 Amino Acid Sequence (domain in AA coordinates:436-501)
MDEETMATGQNRTTVPENLKKHLAVSVRNQWSYGFWSVSASQSGVLEWGDGYNGDIK
TRKTIQASEIKADQLGLRREQLSELYESLSVAESSSSGVAAGSQVTRRASAAALSPEDL
ADTEWYYLVCMFVFNIGEGMPGRTFANGEP IWL CNAHTADSKVFSRSLAKSAAVKT VV
CFPFLGGVVEIGTTEHITEDMNVIQCVKTSFLEAPDPYATILPARSDYHIDNVLDPPQIL
GDEIYAPMFSTPEFPPTASPSRTTNGFDQEHQVADDHDSFMTERTGGASQVQSWQLMDD
ELSNVCVHQSLSNSSDCVSQTFVEGAAGR VAYGARKSRVQRLGQIQEQQRNVKTL SFDPRND
DVHYQSVISTIFKTNHQLILGPQFRNCDKQSSFTRWKSSSSSSSGTATVTAPSQGM LKKI
IFDVPRVHQKEKLM L D S P E A R D E T G N H A V L E K K R E K L N E R F M T L R K I I P S I N K I D K V S I
LDDTIEY L Q E L E R R V Q E L E S C R E S T D T E T R G T M T M K R K K P C D A G E R T S A N C A N N E T G N G K
KVSVMNVGEAEPADTGFTGLTDNLRI G S F G N E V V I E L R C A W R E G V L L E I M D V I S D L H L D S
HSVQSSTGDGLLCLTVNCKHKGSKIATPGMIKEALQ R V A W I C *

>G634 (1..798)
ATGGAGCAAGGAGGAGGTGGTGGTGGTGAATGAAGTTGTGGAGGAAGCTTACCTATTAGT
TCAAGACCTCCTGCTAACAACCTTAGAAGAGCTTATGAGATTCTCAGCCGCCGCGGATGAC
GGTGGATTAGGAGGTGGAGGTGGAGGAGGAGGAGGAAGTGCTTCTTCTTCATCGGGA
AATCGATGGCCGAGAGAAGAACTTTAGCTCTTCTTCGGATCCGATCCGATATGGATTCT
ACTTTTCGTGATGCTACTCTCAAAGCTCCTCTTTGGGAACATGTTTCCAGGAAGCTATTG
GAGTTAGGTTACAAACGAAGTTCAAAGAAATGCAAAGAGAAATTCGAAAACGTT CAGAAA
TATTACAAACGTACTAAAGAACTCGCGGTGGTCTGTCATGATGGTAAAGCTTACAAGTTC
TTCTCTCAGCTTGAAGCTCTCAACACTACTCCTCCTCCTCCTCTCTCATCCTCAGCT
CATCAACCAAGAACAAACAACAACAACCAACAAGAGATGGTCTAGAGCTCGGAA
CAATCATCATTTACCATCATCATCAAGATGGCCAAAGGCAGAGATTCTAGCGCTTATAAAC
CTGAGAAGTGGAATGGAACCAAGGTACCAAGATAATGTACCTAAAGGACTTCTATGGGAA
GAGATCTCAACTTCAATGAAGAGAATGGGATACAACAGAAACGCTAAGAGATGTAAAGAG
AAATGGGAAAACATAAACAATACTACAAGAAAGTTAAAGAAAGCAACAACAGCAACTAC
ACAACAGAATCAATGA
>G634 Amino Acid Sequence (domain in aa coordinates: 62-147, 189-245)
MEQGGGGGGNEVVEEASPISSRPANNLEELMRFSAAADDDGLGGGGGGGGGSSASSSSG
NRWPREETLALLRIRSDMDSTFRDATLKAPLWEHVSRLKLELGYKRSKKCKEKFENVQK
YYKRTKETRGGRHDGKAYKFPFSQLEALNTTPPPPPSHPHAHQPEKQQQQPQQEMVMSSE
QSSLPSSSRWPKAELILALINLRSGMEPRYQDNV PKGLLWEEISTSMKRMGYNRNAKRCKE
KWEININKYKVKESNNSNYMKNQ*

>G676 (1..612)
atgagaaagaaagtaagtagtagtggtgacgaaggaacaatgagtacaagaaaggttg
tggacagtagaagaagacaaaatcctcatggattatgtcaaagctcatggcaaaggtcac
tggaaatcgattatgcaaaaagactggtttaaagagatgtggaaagagttgtagattgagg
tggatgaattatctcagccctaattgtgaaaagaggcaatttcaccgagcaagaaggagat
cttatcattaggtccacaagttgcttggtaataggtggtctttaattgctaaaagagtg
ccgggtcgaacggataatcaagtgagaactattggaacacgcatcttagtaagaaactc
ggaatcaaagatcagaaaacaaacagagcaatggtgatattgtttatcaaataatctc
ccgaatcctaccgaaacatcagaagaagcaaaaatctcgaatattgtcgataacaataat
atcctcggagatgaaattcaagaagatcatcaaggaagtaactacttgagttcactttgg
gttcatgaggatgagtttgagcttagcacactcaccaacatgatggactttatagatgga
cactgtttttga

>G676 Amino Acid Sequence (domain in AA coordinates: 17-119)
MRKKVSSSGDEGNNEYKKGLWTVEEDKILMDYVKAHGKGHWNRIAKKTGLKRCGKSCRLR
WMNYLSPNVKRGNFTEQEEDLIIRLHKLLGNRWSLIAKRVPGRTDNQVKYWNTHLSKKL
GIKDQKTKQSNQDIVYQINLPNPTETSEETKISNIVDNNILGDEIQEDHQGSNYLSSLW
VHEDEFELSTLTNMMDFIDGHCF*

>G682 (1..228)
ATGGATAACCATCGCAGGACTAAGCAACCCAAGACCAACTCCATCGTTACTTCTTCTTCT
GAAGAAGTGAGTAGTCTTGAGTGGGAAGTTGTGAACATGAGTCAAGAAGAAGATTG
GTCTCTCGAATGCATAAGCTTGTCGGTGACAGGTGGGAAGTATAGCTGGGAGGATCCCA
GGAAGAACCCTGGAGAAATTGAGAGGTTTTGGGTATGAAAAATTGA

>G682 Amino Acid Sequence (domain in AA coordinates 27-63)
MDNHRRTKQPKTNSIVTSSSEEVSSLEWVVNMSQEEEDLVSRMHKLVGDRWELIAGRIIP

GRTAGEIERFWVMKN*

>G635 (1..993)

ATGGAGATCATGCGTCCAGGGGTCCTCAGAAAACACTTTGAAAGGAAAAATAAGAATCACA
ACGCGGTGCATGTGGCTTGACAAAGGAAGACTTTTAGATGCACTTCACAAAGCAGCTCAT
GCTGCTCTATCAAGTTGTCTCTGTGACATGTCCCTTGTCTCACATGGAAAGAACAGTCTCC
GAAGTCTTGAGGAAGATTGTAAGGAAGTACAGTGGTAAAAGGCCGTGAAGTCATCGCTATA
GCCACTGAGAATCCAATGGCTGTCCGAGCTGATGAGGTGAGTGCAGAGACTGTCTGGTGAT
CCAAGTGTGGTTCTGGAGTTGCAGCTTTAAGGAAAGTTGTTGAAGGAAATGACAAAAGA
AGTCGGGCGAAGAAAGCACCTTCAACAAGAGCTTCCCCCAAAGAGTAGATCGCACTTTG
GAAGATGATATCATTGATAGTGCAAGACTACTGGCTGAAGAAGAAACTGCGGCATCAACA
TACACGGAAGAAGTTGATACGCCCGTTGGGAGTTCTTCAGAAGAGTCAGACGATTTTGG
AAATCATTATCAATCCATCATCGTCACCTTACCCGAGTGAAACAGAAAATATGAATAAG
GTAGCTGATACGGAGCCTAAAGCAGAGGGTAAGGAAAACAGCAGAGACGACGATGAATTA
GCTGATGCTTCAGATTCTGAAACCAAGTCATCACCAAAACGTGTGAGGAAGAACAATGG
AAACCGGAGGAGATAAAGAAGGTAATCAGAATGCGAGGAGAGCTGCACAGTAGATTTCAA
GTGGTGAAAGGTAGAAATGGCATTGTGGGAAGAGATCTCTTCAAATCTATCAGCTGAAGGA
ATCAATCGAAGCCCGGACAAATGCAAAATCTCTCTGGGCATCACTTATTTCAGAAATACGAG
GAGAGCAAGGCTGATGAGAGAAGCAAGACGAGTTGGCCACATTTTGAGGATATGAACAAC
ATTTTGTGAGAGCTAGGCACACCTGCGTCTTAA

>G635 Amino Acid Sequence (domain in AA coordinates: 239-323)

MEIMRPGVSENTLKGKIRITTRCMWLDKGRLLDALHKAHAALSSCPVTCPLSHMERTVS
EVLKRIVRKYSGRKPEVIALATENPMAVRADSVSARLSGDPVSGVAAALRKVVEGNDKR
SRAKKAPSQEASPKVEDRTLEDDIIDSARLLAEEETAASTYTEVDTPVGSSEESDDFW
KSFINPSSSPSPSETENMNKVADTEPKAEGKENSRRDDELADASDSETKSSPKRVRKNKW
KPEEIKKVIRMRGELHSRFQVVKGRMALWEEISSNLSAEGINRSPGQCKSLWASLIQKYE
ESKADERSKTSWPHFEDMNNILSELGTPAS*

>G1068 (150..1310)

GAGAGTTGTTAGCTAGCTCACACGCTTTCGCTTAAAACTCAAAAACCTGCACCTTCTCGT
CTATTTTCTCGGCATTTCGTAACAGAAAAGTGGGTCTCCAAGAAAATTACCCCTAAATTC
ACAAAGATTCTACTTTTCTCCACCTCCAATGGATTCCAGAGAGATCCACCACCAACAAC
AGCAACAACAACAACAACAACAGCAGCAGCAGCAACAACAGCAACATCTACAACAACAGC
AACAACCACCGCCAGGGATGTTAATGAGTCACCACAATTCTTACAATCGAAACCTAACG
CCGCCGCCGCTGTTTTAATGGGTCAACACCTCCACATCTCAAGCTATGCATCAAAGAT
TACCTTTTGGTGGTTCTATGTACCCGCATCAGCCTCAACAACATCAGTATCATCATCTC
AGCCTCAGCAACAGATAGATCAGAAGACTCTTGAATCTCTTGGATTCTTACTTTCGCTC
TTCCTTCTGCTTCTAATCTTACGGTGGTGGAAATGAAGGAGGTGGTGGTGGTATAGCG
CCGGAGCTAATGCTAACTCTTCCGATCCACCTGCTAAACGGAACAGAGGACGTCCTCTG
GCTCCGGTAAGAAGCAGCTCGATGCTTTAGGAGGAACAGGAGGAGTTGGGTTACGCCCTC
ATGTCATTGAGGTTAAACAGGAGAGGACATAGCTACGAAGATATTGGCGTTTACGAACC
AAGGGCCACGCGCAATCTGTATTCTCTCAGCTACAGGAGCTGTAACATAATGTGATGCTTC
GTCAAGCTAACAATAGCAATCTTACTGGAAGTGTAAAGTATGAGGGCCGATTTGAAATCA
TTTCTCTGTGAGTTCTTTCTTGAATTCTGAGAGTAATGGTACTGTGACCAAACTGGTA
ACTTGAGTGTGTGCTGGCTGGACACGAAGGCCGATTGTGGGTGGATGTGTTGATGGAA
TGCTAGTAGCTGGATCACAAGTCCAGGTCAATTGTGGGAAGCTTTGTACCAGATGGAAGGA
AGCAGAAACAAAGTGGGGGCGTGCTCAGAATACTCCGGAGCCAGCTTCAGCACCAGCCA
ATATGTTGAGCTTTGGTGGTGTGGTGGACCGGAAGCCCTCGATCTCAAGGACAACAAC
ACTCGAGCGAGTCATCAGAGGAAAACGAAAGTAATTCTCCGTTGCACCGTAGAAGCAACA
ACAACAACAGCAACAATCATGGGATATTTGGAACTCTACACCTCAACCGCTTCACCAAA
TTCCTATGCAGATGTACCAGAATCTCTGGCCTGGCAACAGTCTCAATAAACAGATGGTT
CATGGGTCAAGATTTGACCGGGTTGCTTCTCTGTTCTTTTGACACATCTCTCCATCAG
ATTTATCTCTATAAAGTAGATTGAGCTCTCTTACTCTCTCATCTTCTCTCTTACTAT
TTCTCTTAAATTTAGCTTTGGTTTGTAGATAAATAGAGAGAGAGACATGTTAAGTAGGT
TTCAAATTCATCTTGTGTTAGTTTGTCTTAGTAGTTCTTTTGATTGTGATGATCATA
AAGACTTGTCTTTTCTCTCTATATTCAACGAATTATCCACTTTAA

>G1068 Amino Acid Sequence (domain in AA coordinates: 143-150)

MDSREIHQQQQQQQQQQQQQQQHLQQQQQPPPGMLSHHNSYNRNPNAAAVLMGHN
TSTSQAMHQRLFPGGSSPHQPQQHQYHHPQPQQIDQKTLES LGFPTSPLPSASNSYGG

GNEGGGGDSAGANANSSDPPAKRNRGRPPGSGKKQLDALGGTGGVGFTHPHVIEVKTGED
IATKILAFNTNQGPRAICILSATGAVTNVMLRQANNSNPTGTVKYEGRFEIISLSGSFLNS
ESNGTVTKTNLSVSLAGHEGRIVGGCVDGMLVAGSQVQVIVGSFVPDGRKQKQSAGRAQ
NTPEPASAPANMLSFGGVGGPGSPRSQGGQHSSESSEENESNSPLHRRSNNNNNSNNHGIF
GNSTPQLHLQIPMQMYQNLWPGNSPQ*

>G1225 (1..984)

ATGACTCTAGAAGCTTTATCATCAAACGGTCTTTTAACTTTTGTCTCTGAAACTCTT
TCACCAACTCCATTCAAGTCTCTCGTCGATCTCGAGCCATTGCCGGAATGATGTCATC
ATATCGAAGAACAGAATTTTCGGAGATATCTAATCAAGAACCACCACAGCGACAACCA
CCAGCTACGAATCGAGGGAAGAAGCGGCGGAGGAGGAAGCCTAGGGTTTGCAAAAACGAG
GAAGAAGCTGAGAATCAACGAATGACTCACATTGCCGTCGAAAGAAATCGAAGAAGACAA
ATGAATCAACATCTCTCTGTCTTGGCCTCTCATGCTCAACCTTTTGTCTCACAAGGT
GATCAAGCTTCAATAGTTGGTGGAGCCATAGATTTTATCAAAGAACTTGAACACAAATTA
CTATCTCTTGAAGCTCAAAAACATCATAATGCTAAATTAACCAGTCGGTTACTTCTTCA
ACAAGTCAAGACTCAATGGTGAACAAGAGAATCCTCATCAACCATCTTCACTATCTCTA
TCGCAGTTCTTTCTTCAATCATAACGATCCGAGCCAAGAGAATAGGAACGGCTCAACAAGC
TCGGTGAACCCCTATGGAAGATCTTGAGGTGACTCTAATCGAAACTCATGCTAACATC
AGAATCTTGTGTCGAGAAGAAGAGGTTTCCGGTGGAGCAGCTTGGCCACCACCAACCGCCG
CAGCTTTCGAAGCTGGTGGCTTCTCTACAATCGCTGTCCCTCTCCATTCTTCACCTTAGT
GTCAACAACATTGGACAATTATGCTATTTACTCCATCAGCGCTAAGGTGGAAGAGAGTTGC
CAGCTAAGTTCAGTAGATGACATTGCAGGAGCAGTTCAACCATGCTAAGTATCATTGAA
GAGGAGCCTTTTTGTTGCTCATCAATGTGAGAATTAACATTTGACTTCTCTTTGAATCAC
TCAATGTCACTCATCTCTCTGAGAAATCTCTTTTTTGTGTTGTTATTCCTTCTTTTA
ATTTTATCACATAGCACATCTTTAGTTTTTTTTTTT

>G1225 Amino Acid Sequence (domain in AA coordinates: 78-147)

MTLEALSSNGLLNFLSETLSPTPFKSLVDLEPLPENDVLIISKNTISEISNQEPQPQRQP
PATNRGKKRRRRKPRVCKNEEEAENQRMTHIAVERNRRRQMNQHLVLSLMPQPFQAHKG
DQASIVGGAIDFIKELEHKLSSLEAQKHNAKLNSVTSSTSDSNGEQENPHQPSLSL
SQFFLHSYDPSQENRNGSTSSVKTPMEDLEVTLIETHANIRILSRRRGFRWSTLATTKPP
QLSKLVASLQSLSLILHLSVTLTDNYAIYSISAKVEESCQLSSVDDIAGAVHMLSIIE
EEPFCCSSMSELPFDLSLHNSNVTHSL*

>G1337 (97..1398)

AATGGATTGTGCATCATCTCTCACCGTCTCTAGTCTCTGAAAATAAATTCTGATTTTG
ATTTTCAATTTTAGGGATTTTGAGAGAGAGTCAGTTATGAGTAGTTCGGAGAGAGTACCG
TGCGATTTTTCGGCGAGCGTACGGCGGTTTGTGTTTGTAGAGCCGATACGGCGAAGCTG
TGTTTGCCTTGTGATCAGCAAGTTCACACGGCGAATCTGTTGTCGAGGAAGCACGTGCGA
TCTCAGATCTGCGATAATTGCGGTAACGAGCCAGTCTCTGTTCCGGTGTTCACCGATAAT
CTGATTTTGTGTCAGGAGTGTGATTGGGATGTTACCGGAAGTGTTCAGTTTCCGATGCT
CATGTTTCGATCCGCCGTGGAAGGTTTTTCCGGTGTCCATCCGGCTTGGAGCTTGCTGCT
TTATGGGGACTTGATTTGGAGCAAGGGAGGAAAGATGAAGAGAATCAAGTTCGATGATG
GCGATGATGATGATAATTCGGGATGCAGTTGGATTCTTGGGTTTTGGGATCTAATGAA
TTGATTGTTCCAGCGATACGACGTTTAAAGACGCTGGATCTTGTGGATCTAGTTGTGGG
AGGTATAAGCAGGTATTGTGTAAGCAGCTTGAGGAGTTGCTTAAGAGTGGTGTGTCGGT
GGTGTGCGGATGATGGTGTGATCGTGACCGTGATTGTGACCGTGAGGGTGCTTGTGATGGA
GATGGAGATGGAGAAGCAGGAGAGGGGCTTATGGTTCCGGAGATGTCAGAGAGATTGAAA
TGGTCAAGAGATGTTGAGGAGATCAATGGTGGCGGAGGAGGAGGTTAACCAGCAGTGG
AATGCTACTACTACTAATCCTAGTGGTGGCCAGAGTTCTCAGATATGGGATTTTAACTTG
GGACAGTCACGGGGACCTGAGGATACGAGTCGAGTGGAAGCTGCATATGTAGGGAAAGGT
GCTGCTTCTTCAATCAACATCAACAATTTTGTGACCATATGAATGAACTTGTTCCTACT
AATGTGAAAGGTGTCAAAGAGATTAAGAAAGGATGACTACAAGCGATCAACTTCAGGCCAG
GTACAACCAACAAATCTGAGAGCAACAATCGTCAATTACCTTTGGCTCTGAGAAAGGT
TCGAACTCCTCCAGTGACTTGCACTAAGGCTGATCTGGAGCGGCTGGCTCAGAACAGAGGAGATGCA
ATGCACCGTTACAAGGAAAAGAGGAAGACACGGAGATATGATAAGACCATAAGGTATGAA
TCGAGGAAGGCAAGAGCTGACACTAGGTTGCGTGTGAGGCGAGATTTGTGAAAGCTAGT
GAAGCTCCTTACCCTTAACCTTAAGTTTTTTCACATAGGCTTCTTTTAGCTACAACTT
AGTTACTTTTTTACTCCACTGCCCTCATAAATGTACAGACCGGTCTCGTTTCATCTGGCC

GCCCTTCTTGTTTTATTGCCTTATCTGGCCCTTTTATGTACCTTGAATCTTATCTAGTT
TAAAAAAGATTGTAACCTTCTAGAAAACCATATTCTGTTGACAGTATATACATGTCTATC
CAAGCAAAAA

>G1337 Amino Acid Sequence (domain in AA coordinates: 9-75)
MSSSERVPCDFCGERTAVLFCRADTAKLCLPCDQQVHTANLLSRKHVRSQICDNCNEPV
SVRCFTDNLILCQECDWDVHGSCSVSDAHVRSVEGFSGCPSALELAALWGLDLEQGRKD
EENQVPMAMMDNFMQLDSWVLGSNELIVPSDTTFKKRGSCGSSCGRYKQVLCKQLEE
LLKSGVVGGDGDGDRDRDCDREGACDGDGDGEAGEGLMVPESERLKWSRDVEEINGGG
GGGVNQWNATTNPSSGGQSSIWDFNLGQSRGPEDTSRVEAAYVGKGAASSFTINNFDV
HMNETCSTNVKGVKEIKKDDYKRSTSGQVQPTKSESNNRPITFGSEKGSNSSDLHFTEH
IAGTSCKTTRLVATKADLERLAQNRGDAMQRYKEKRKTRRYDKTIRYESRKARADTRLRV
RGRFVKASEAPYP*

>G1759 (110..700)
CGAGAAAAGGAAAAAAAAAATAGAAAGAGAAAACGCTTAGTATCTCCGGCGACTTGAAC
CCAAACCTGAGGATCAAATTAGGGCACAAAGCCCTCTCGGAGAGAAGCCATGGGAAGAAA
AAACTAGAAATCAAGCGAATTGAGAACAAAAGTAGCCGACAAGTCACCTTCTCCAAACG
TCGCAACGGTCTCATCGAGAAAGCTCGTCAGCTTTCTGTTCTCTGTGACGCATCCGTCGC
TCTTCTCGTCTCTCCGCTCCGGCAAGCTCTACAGCTTCTCCTCCGGCGATAACCTGGT
CAAGATCCTTGATCGATATGGGAAACAGCATGCTGATGATCTTAAAGCCTTGATCATCA
GTCAAAGCTCTGAACATATGGTTCACACTATGAGCTACTTGAACCTTGATGATGCAAGCT
TGTGGGATCAAATGTCAAAGATGTGAGTATCGATGCTCTTGTTCACCTGGAGGAACACCT
TGAGACTGCCCTCTCCGTGACTAGAGCCAAGAAGACCGAACTCATGTTGAAGCTTGTGTA
GAATCTTAAAGAAAAGGAGAAAATGCTGAAAGAAGAGAACCAGGTTTTGGCTAGCCAGAT
GGAGAATAATCATCATGTGGGAGCAGAAGCTGAGATGGAGATGTCACCTGCTGGACAAAT
CTCCGACAATCTTCCGGTGACTCTCCCACTACTTAATTAGCCACCTTAAATCGGCGGTTG
AAATCAAAATCCAAACATATATAATTATGAAGAAAAAAAAAATAAGATATGTAATTATT
CCGCTGATAAAGGCGAGCGTTTGTATATCTTAATACTCTCTCTTTGGCCAAGAGACTTTG
TGTGTGATACTTAAGTAGACGGAACCTAAGTCAATACTATCTGTTTTAAGACAAAAGGTTG
ATGAACCTTTGTACCTTATTCGTGTGAGAAAAAAAAAAAAAAAAA
>G1759 Amino Acid Sequence (conserved domain in AA coordinates: 2-57)
MGRKKLEIKRIENKSSRQVTFKRRNGLIEKARQLSVLCDASVALLVVSASGKLYSFSSG
DNLVKILDYRKQHADDLKDLDHQSALNYGSHYELLELVDSKLVGSNVKNVSDALVQL
EEHLETALSVTRAKKTEMLKLVENLKEKEKMLKEENQVLASQMNHNHVGAEAEEMESP
AGQISDNLPLVTLPLLN*

>G1804 (169..1497)
TATCTCTCTCTTTCTCAAAACCTTTTCAGTCAAAATTCTCCGGCGGCTTTTAAACTATGTG
AAGGAGGAGAACCTCCATAACAAGAAGCGGATTCTCTCAGTTTTCGGCGGCGGAGGAAC
ACAAAGCCACCGGTTTTTAGACACACAGATTTCATTTTTCAGTTGTTAAATGGTAACTAGA
GAAACGAAGTTGACCTCAGAGCGAGAAGTAGAGTCGTCCATGGCGCAAGCGAGACATAAT
GGAGGAGGTGGTGGTGAGAATCATCCGTTACTTCTTTGGGAAGACAATCCTCTATCTAC
TCATTGACCTTTGACGAGTTCCAACATGCTTTATGTGAGAACGGCAAGAAGCTTTGGGTCC
ATGAACATGGACGAGTTTCTTGTCTCTATTTGGAACGCAGAGGAGAATAATAACAATCAA
CAACAAGCAGCAGCAGCTGCAGGTTACATTCTGTTCCGGCTAATCACAATGGTTTCAAC
AACAACAATAACAATGGAGGCGAGGGTGGTGGTGGTGTCTTTAGTGGTGGTTCTAGAGGC
AACGAAGATGCTAACAATAAGAGAGGGATAGCGAACGAGTCTAGTCTTCTCGACAAGGC
TCTTTGACACTTCCAGCTCCGCTTTGTAGGAAGACTGTTGATGAGGTTTGGTCTGAGATA
CATAGAGGTGGTGGTAGCGGTAATGGAGGAGACAGCAATGGACGTTAGTAGTAGTAAT
GGACAGAACAATGCTCAGAACGGCGGTGAGACTGCGGCTAGACAACCGACTTTTGGAGAG
ATGACACTTGAGGATTTCTTGGTGAAGGCTGGTGTGGTTAGAGAATCCCACTAATCCT
AAACCTAATCCAAACCCGAACCAAAACCAAAACCGTCTAGTGTAATACCCGAGCTGCA
CAGCAACAGCTTTATGGTGTGTTCAAGGAACCGGTGATCCTTCATTCCCGGTCAAGCT
ATGGGTGTGGGTGACCCATCAGGTTATGCTAAAAGGACAGGAGGAGGAGGATCAGCAG
GCGCCACAGGTTGAGGAGGTTGCTATGGAGGTGGCGTTGGGTTTGGAGCGGGTGGGA
CAGCAATGGGAATGGTTGGACCGTTAAGCCCGGTGCTTTCAGATGGATTAGGACATGGA
CAAGTGGATAACATAGGAGGTGAGTATGGAGTAGATATGGGAGGGCTAAGGGGAAGGAAA
AGAGTAGTGGATGGTCCAGTGGAGAAAGTAGTGGAGAGAAGACAGAGGAGGATGATCAAG
AACCGCGAGTCTGCTGCTAGATCTAGAGCAAGAAAACAAGCATATACAGTGAATTGGA

GCTGAACCTTAACCAAGTTGAAAGAAGAGAATGCGCAGCTAAAAATGCATTGGCGGAGTTG
GAGAGGAAGAGGAAGCAACAGTATTTTGAAGTCAAGGGCACAACCGAAATTG
CCGAAATCGAACGGGAGATTGCGGACATTGATGAGGAACCGAGTTGTCCACTCTAAACA
AACAAATAGGAAGATGGAGAAGAAGTCGGAGACAGAACGAGGGAAAACTGATGATTTTCT
ACGTTGTTGTTTTGCTTTTGAAGGAATGAGGTTATAGAATCTTTATACTTTGATGTTTTCT
GTGTTGGTAGGAGGAACACCATCTGATCTGCTTTACTAGTGTCCCTGTGAACAAAGAAA
GTGATTCTGTGTTTCAACATCATCAATCTTTGGAAA

>G1804 Amino Acid Sequence (domain in AA coordinates: 357-407)

MVTRETKLTSEREVESMAQARHNGGGGGENHPFTSLGRQSSIYSLTLDEFQHALCENGK
NFGSMNMDEFVLVSIWNAEENNNNQQAAGASHSVPAHNGFNNNNNNGGEGVGVFSG
SSRGNEDANNKRGIANESSLPQGSITLPAPLCRKTVDEVWSEIHRGGSGNGGDSNGRS
SSSNGQNNNAQNGGETAARQPTFGEMTLEDFLVKAGVREHPTNPKPNPNPNQNPSSVI
PAAQQQLYGVFQGTGDPSPFGQAMGVGDPSPGYAKRTGGGGYQQAPPVQAGVCYGGGVGF
GAGGQQMGMVGPLSPVSSDGLGHGQVDNIGGQYGVDMGGLRGRKRVRVDGPVEKVVERRQR
RMIKNRESAARSARKQAYTVELEAELNQLKEENAQLKHALAELERKRKQYFESLSRA
QPKLPKSNGLRLRTLNRNPSCPL*

>G207 (16..930)

aaaagatctgtttcaatggcggatcgtgttaaaggtccatggagtcagaagaagatgag
cagctacgaaggatggttgagaaatacggaccgaggaattggtctgcgattagcaaatacg
attccaggtcgatctggtaaatcgtgtagattacgttggtgtaatacagttatctccggag
gttgagcatcgtcctttctcgcggaggaagatgagactattgtaaccgcccgtgctcag
tttggttaacaagtggtggcgacgattgctcgtcttcttaacggctcgtagcgataacgccgtt
aaaaatcactggaactctacgcttaagaggaatgcagcggaggtgtggcggttacgacg
gtgacggagacggaggaagatcaggatcggccgaagaagaggagatctgttagctttgat
cctgcttttgctccggtggatactggattgtacatgagtcctgagagtcctaaccggaatc
gatgttagtgattcttagcacgattccgctcaccgctcgtctcctggtgctcagctgtttaaa
ccatgcccgaatttccggcggttttacgggtggttccgcagccggttaccggttgaaatgtct
tcgtcttcggaggatccacctaacttcgttgagtttgcactacctggagctgagaacacg
agttcgagccataacaataacaacaacgcgttgatgtttccgagatttgagagtcagatg
aagattaatgtagaggagagaggaggaggaggaaggacgtagaggtgagtttatgacg
gtggtgcaggagatgataaaaagctgaagtgaggagttacatggcgaaatgcagaaaaca
agtgggtggttctcgtcgtcggaggtttatacgaatccggcggcaatggtggttttagggat
tgtggagtaataacacctaaggttgagtagttttggtttagggttaaaacttgaatcgat
tggtggttttcaagagcattcatttttgggttttatggtaaaattaaaaacaaaaacaaa
atgtacagaggaattaaaatttctatggaataatcttaaatctcaaatatttggtacttg
tttgggtgattcataacccaaatcaaa

>G207 Amino Acid Sequence (domain in AA coordinates: 6-106)

MADRVKGPWSQEEDQLRRMVEKYGPRNWSAISKSIPIGRSGKSCRLRWCNQLSPEVEHRP
FSPEEDETIVTARAQFGNKWATIIARLLNGRTDNAVKNHWNSTLKRKCSGGVAVTTVTETE
EDQDRPKKRRSVSFDPAFAPVDITGLYMSPEPNDIDVSDSTIPSPSSPVAQLFKPMPIS
GGFTVVPQPLPVEMSSSEDPPSTLSLSLPGAENTSSSHNNNNNNALMFPRFESQMKINVE
ERGGGGEGRRGFEFMTTVQEMIKAEVRSYMAEMQKTSGGFVVGGLYESGGNGGFRDCGVIT
PKVE*

>G218 (1..1182)

ATGGAGGCAGAGATCGTGAGACGATCGGAGGTAACGGGATTAAGAAGGGAGGTGGAAGAA
TCGTCAATTGGTAGAGGAGATTGCGATGGTGATGGCGGCGATGTGGGAGAAGATGCGGCA
GGGTTGCTTGGGACGAGCGGGAGAGGAAGAAGAGATCGAGTTAAAGGGCCGTGGTCGAAG
GAGGAGGATGATGTGTTGAGTGAGCTCGTTAAGAGGTTGGGAGCGAGGAATTGGAGTTTT
ATCGCTCGGAGTATTCCTGGTTCGTTTCAGGCAAGTCTTGTCGTTCTCGTTGGTGTAATCAG
CTCAATCCAAATCTTATACGCAATTCATTTACTGAGGTAGAGGATCAGGCTATCATCGCA
GCACATGCCATCCACGAAACAAATGGGCTGTTATCGCGAAGCTCCTCCCGGAAGAACA
GATAATGCTATCAAGAACCACTGGAACCTCTGCTTTAAGACGTCGATTCATAGACTTTGAA
AAGGCCAAGAAATAGGAACCTGGAAGCTTGGTCTGGATGATCTGGATTTGACAGAACG
ACAACAGTAGCCTCATCAGAAGAACTTTATCTTCAGGCGGTGGTTGCCATGTAACACT
CCAATTGTATCTCCAGAAGGCAAAGAAGCTACCACCTCCATGGAAATGTCTGAAGAACAA
TGCGTAGAGAAAACAAACGGAGAAGGTATTTCTAGGCAAGATGATAAGGATCCTCCACG
CTTTTCCGCCCCAGTGCCTCGGCTCAGTTCTTTAATGCTTGCAATCACATGGAAGGATCA

CCCTCTCCACATATACAAGACCAAAATCAGCTCCAATCATCTAAACAAGACGCAGCAATG
CTAAGATTGCTTGAAGGAGCTTACAGCGAACGGTTTGTGCCTCAAACATGTGGAGGTGGT
TGTTGCAGCAACAATCCCGATGGCAGTTCCTCAGCAAGAATCATTGTTGGGTCCAGAGTTT
GTGGATTACTTAGACTCACCAACGTTTCCGAGTTCGGAAGTCTGCTATAGCAACGGAA
ATAGGCAGCCTCGCTTGGCTGAGAAGCGGTTTAGAGAGTAGCAGCGTGAGGGTGATGGAA
GACGCAGTTGGTCGGTTAAGGCCTCAAGGCTCCAGGGGTATCGAGATCATTATCTTGTA
TCTGAACAGGGGACGAACATAACCAATGTCCTGTCCACATAA

>G218 Amino Acid Sequence (domain in AA coordinates: TBD)

MEAEIVRRSEVTGLRREVEESSIGRGDCDGDGVDGEDAAGFVCTSGRGRDRVKGPWSK
EEDDLVSELVKRLGARNWSFIARSIPGRSGKSCRLRWCNQLNPNLIRNSFTEVEDQAIIA
AHAIHGNKWAIVIAKLLPGRTDNAIKNHWSALRRRFIDFEKAKNIGTGLVVDSDGFDRT
TTVASSEETLSSGGGCHVTTPIVSPEGKEATSMEMSEEQCVEKTNGEGISRQDDKDPPT
LFRPVPRLSSFNACNHMEGSPSPHIQDQNLQSSKQDAAMLRLLEGAYSERFVPQTCCGG
CCSNPDGFSFQQESLLGPEFVDYLDSPTFPSSELAAIATEIGSLAWLRSGLESSSVRVME
DAVGRLRPQSGRHRDHYLVSEQGTNITNVLST*

>G241 (46..867)

GAAAAACATTTCACTTCTTTTATCAGCAATCACAATCAAAGAGATGGGAAGAGCTCCA
TGCTGTGAGAAGATGGGGTTGAAGAGAGGACCATGGACACCTGAAGAAGATCAAATCTTG
GTCTCTTTTATCCTCAACCATGGACATAGTAAGTGGCGAGCCCTCCCTAAGCAAGCTGGT
CTTTTGAGATGTGGAAGAGCTGTAGACTTAGGTGGATGAACATTTTAAAGCCTGATATT
AAACGTGGCAATTTACCAAAGAAGAGGAAGATGCTATCATCAGCTTACACCAAATACTT
GGCAATAGATGGTCAGCGATTGCAGCAAACTGCCTGGAAGAACCATAACGAGATCAAG
AACGTATGGCACACTCACTTGAAGAAGAGACTCGAAGATTATCAACAGCTAAACCTAAG
ACCAGCAACAAAAGAGGGGTACTAAACCAAATCTGAATCCGTAATAACGAGCTCGAAC
AGTACTAGAACGGAATCGGAGCTAGCAGATTTCATCAAACCTTCTGGAGAAAGCTTATTT
TCGACATCGCCTTCGACAAGTGAGGTTTCTTCGATGACACTCATAAGCCACGACGGCTAT
AGCAACGAGATTAATATGGATAACAAACCGGGAGATATCAGTACTATCGATCAAGAATGT
GTTTCTTTTCGAACTTTTGGTGCGGATATCGATGAAAGCTTCTGGAAAGAGACTGTAT
AGCCAAGATGAACACAACTACGTATCGAATGACCTAGAAGTCGCTGGTTTAGTTGAGATA
CAACAAGAGTTTCAAACCTTGGGCTCCGCTAATAATGAGATGATTTTTCAGTGTAGATG
GAACTTCTGGTTCGATGTATTGGCTAGAACCGGCGGGGAACAAGATCTCTTAGCCGGGCT
CTAGTTAACATGTTTGAGGAGTAAAGTGAAATGGTGCAAATTAGTTAAGGCTAAGAAATT
CAAAAGCTTTTGTGTTTACCAGAGAAAAACACACTCTAACTCTTGATGTGATGTAGTTAGT
GTATTAATTAGAGGCTGCGTTTTCAA

>G241 Amino Acid Sequence (domain in AA coordinates: 14-114)

MGRAPCCEKMGLKRPWTPFEEDQILVSFILNHGHSNWRALPKQAGLLRCGKSCRLRWMNY
LKPDIKRGNFTKEEEDAIISLHQILGNRWSAIAKLPGRITDNEIKNVVHHLKRLLEDYQ
PAKPKTSNKKKGTKPKSESIVITSSNSTRSESELADSSNPSGESLFTSPSTSEVSSMTLI
SHDGYSNEINMDNKPGLDISTIDQECVSFETFGADIDESFWKETLYSQDEHNYVNSNDLEVA
GLVEIQQEFQNLGSANNEMIFDSEMELLVRCIG*

>G254 (15..923)

CGATTTTCGAGCTCTATGGTGTCCGTAAACCTAGACCTAAGGGTTTTCCAGTTTTCGATT
CCTCGAATATGAGTTTACCAAGCTCCGATGGATTGGTTTCGATTCCGGCCACGGGACGGA
CCAGTACGGTGTCTGTTTTCTGAGGATCCGACGACGAAGATTCCGAAGCCGTACACAATCA
AGAAGTCGAGAGAGAATTGGACAGATCAAGAGCACGATAAATTTCTAGAAGCTCTTCACT
TATTCGATAGGGATTGGAAGAAAATAGAAGCCTTTGTTGGATCAAAAACAGTAGTTCAGA
TACGAAGCCACGCTCAGAAATACTTTCTCAAAGTTCAGAAGAGTGGTGTACCAACATC
TTCCACTTCTCGACCTAAGAGGAAAGCGAGTCATCCTTATCCTATAAAGGCTCCTAAAA
ATGTTGCTTATACCTCTCTCCCGTCTTCGAGTACATTACCGTTGCTTGAGCCTGGTTATT
TGATATAGCTCTGATTTCGAAGTCATTGATGGGAAACCAAGCTGTTTGTGCATCTACCTCTT
CTTCGTGGAATCATGAATCGACAAATCTGCCAAAACCGGTGATTGAAGAGGAACCGGGAG
TCTCGGCCACGGCTCTCTCCCAAATAATCGCTGCAGACAGGAAGATACAGAGAGGGTAC
GAGCAGTGACAAAGCCAAATAACGAAGAAAGTTGTGAAAAGCCACATAGAGTGATGCCGA
ATTTTGCTGAAGTTTACAGCTTCATTGGAAGTGTCTTCGATCCCAACACATCAGGCCACC
TCCAGAGATTAAAGCAGATGGATCCAATAAATATGGAAACGGTCTTTTACTGATGCAAA
ACCTGTCTGTAAATCTGACAAGTCCCGAGTTTGACAGACAAAGGAGGTTGATATCATCAT
ACAGCGCTAAAGCTTTGAAATAGAGATAGAATAAAACAATAATGTACCTTATGTGAGATC

AAGAGACAATCATCCAAGGTCTGTATGCATTGCTTGGATTTAGGCCTCGTGTTCCTCACTA
CAGGAGCAGAACCAATCGCAAAGACTCTTAGATGGCTACTGAGTTGTGGTTTTTATGTCT
CTGTAAGTCGCGGTGGAGCACACGTGTTTGTCTGTCTTGTGTATGTGTGTATAGATAAT
ACAAGGTTTTGTCAGAGTAAGGTCACAGTTAGCTGCAAGTGAGTTTGGATCAATCTTAAGA
TTAAAACCTGAGAGTGAGTGTCCAAAGAGACTGTGTAATATTGGTTTGGCGGTGAGCAG
AAGAGTTTTGAAGTGCACATCCAGTTAGTGATAACACGGTTGAAGAAAAGGTAAGGTTAC
AAGTTTAGTTTTGAATAATTGTATACTCAAAAAATATGAATGTATAAAGAATAATCACTT
GAGTCGCCTTA

>G254 Amino Acid Sequence (domain in AA coordinates: 62-106)
MVSVNPRPKGFPVFDSSNMSLPSSDGFSGIPATGRTSTVSFSEDPPTTKIRKPYTIKKSRE
NWTQEHDKFLEALHLFDRDWKKIEAFVGSKTVVQIRSHAQKYFLKVQKSGANEHLPLPR
PKRKASHPYPIKAPKNVAYTSLPSSSTLPLLEPGYLYSSDSKSLMGNQAVCASTSSSWNH
ESTNLPKPVIEEPGVSATAPLPNNRCRQEDTERVRAVTKPNNEESCEKPHRVMPNFAEV
YSFTIGSVFDPNTSGHLQRLKQMDPINMETVLLLMQNLSVNLTSPEFAEQRLISSYSAKA
LK*

>G26 (73..729)
TTGGCTTGTAACCAACCCATCTTTGACTTCAAAAATAAAATAAAAATAATCATAATTGA
CATCATCGGATAATGCATAGCGGGAAGAGACCTCTATCACCAGAATCAATGGCCGGAAAT
AGAGAAGAGAAAAAAGAGTTGTGTTGTTGCTCAACTTTGTGCGGAATCTGATGTGTCTGAT
TTTGTCTCTGAACCTACTGGTCAACCCATCCCATCATCCATTGATGATCAATCTTCGTCG
CTTACTCTTCAAGAAAAAAGTAACTCGAGGCAACGAACTACAGAGGCGTGAGGCAAAGA
CCGTGGGGAATGGGCGGCTGAGATTTCGTGACCCGAACAAGGCAGCTCGTGTGTGGCTT
GGGACGTTTCGACACTGCAGAAGAAGCCGCCCTTAGCGTATGATAAAGCTGCATTTGAGTTT
AGAGGTCACAAGGCCAAGCTTAACTTCCCGGAGCATATTCGTGTCAACCCCTACTCAACTC
TATCCATCGCCCGCTACTTCCCATGATCGCATTATCGTGACACCACCTAGTCCACCTCCA
CCAATTGCTCCTTGACATACTTCTTGATCAATATGGCCACTTTCAATCTCGAAGTAGTGAT
TCCAGTGCCCAACTTGTCCATGAATATGCTGTCTTCTTCGTCTTCATCTTTGAATCATCAA
GGGCTAAGACCAAATTTGGAGGATGGTGAACCGTGAAGAACATTAGTATCCACAAACGA
CGAAAATAACATGTTAATGGCATAAATATCTCTTCGTCCAAGTTATCAAACGCATTGACC
TCCGGCTTTGATCATTTTTAGGCGCTTAATCTCTTTACGACTTCATTTTGGTAGTCTTTAA
AGAGTCTATGGAGTGGATTTAGCTAGGAATCAGGCCTTATGGATGAAAAATATATAAATT
TTGAACATGACTATGCAAGAATGGGATGAAGACTACTTAGCTTGGAACCGTCCGTGATAG
GTCATGACGACTATATCCACAGAAGATGACCGACGGAGACAACAACATGCCTCACCTGAT
CGACCGATCAAATGAGATAATGTGTTGACCGGACCGGTCCGATCAGGTTGGGTGAGTAT
ATCA

>G26 Amino Acid Sequence (domain in AA coordinates: 67-134)
MHSGKRPLSPESMAGNREEKELCCCSTLSESDVSDVFVSELTGQPISSIDDQSSSLTLQ
EKSNSRQRYNRGVRQRPWGKWAABIRDPNKAARVWLGTDFDTEEAALAYDKAAFEFRGHK
AKLNFPEHILRVNPTQLYSPATSHDRIIVTPPSPPPPIAPDILLDQYGHFQSRSSDSSAN
LSMNMSSSSSSSLNHQGLRPNLEDEGENVKNISIHKRRK*

>G263 (48..902)
TTTTTAGTTTATTTTTCTGTGGTAAAAATAAAAAAGTTCCGCCGAGATGACGGCTGTGA
CGGCGGCGCAAAGATCAGTTCCGGCGCCGTTTTTAAGCAAACGTATCAGCTAGTTGATG
ATCATAGCACAGACGACGTCGTTTCATGGAACGAAGAAGGAACAGCTTTTGTCTGTGGA
AAACAGCAGAGTTTGCTAAAGATCTTCTTCTCAATACTTCAAGCATAATAATTCTCAA
GCTTCATTCGTGAGCTCAACACTTACGGATTTTCGTAAACTGTACCGGATAAATGGGAAT
TTGCAAACGATTATTTCCGGAGAGGCGGGGAGGATCTGTTGACGGACATACGACGGCGTA
AATCGGTGATTGCTTCAACGGCGGGGAAATGTGTTGTGTTGGTTTCGCCTTCTGAGTCTA
ATTCTGGTGGTGGTGATGATCACGGTTCAAGCTCCACGTCATCACCCGGTTTCGTGGAAGA
ATCCTGGTTTCGGTGGAGAACATGGTTGCTGATTTATCAGGAGAGAACGAGAAGCTTAAAC
GTGAAAACAATAACTTTGAGCTCGGAGCTCGCGGCGGCGAAGAAGCAGCGCATGAGCTAG
TGACGTTCTTGACGGGTCATCTGAAAGTAAGACCGGAACAAATCGATAAAATGATCAAAG
GAGGGAAATTTAAACCGGTGGAGTCTGACGAAGAGAGTGAGTGCGAAGGTTGCGACGGCG
GCGGAGGAGCAGAGGAGGGGTAGGTGAAGGATTGAAATTGTTTGGGGTGTGGTTGAAAG
GAGAGAGAAAAAGAGGGACCGGGATGAAAAGAATTATGTGGTGGTGGTCCCGTATGA
CGGAAATAAAGACGTGGACTTTACGCGCCGTTGTGGAAGCAGCAAAGTCTGCAACT
AAAAAAGAGTAGAAGACTGTTCAAACAGCGTGTGACACGTCATCGACGACGAGAAAA

AAATGATTTAAAAAACTATTTTTTCCGTAAGGAAGAAAAGTTATTTTTATGTTTTAAAA
AGGTGAAGAAGGTCCAGAAGGATCAACGCAATATATAAATGGATTTTCATGTATTATAT
AATTTAATTAGTGTATTAAGAAAA

>G263 Amino Acid Sequence (domain in AA coordinates: TBD)

MTAVTAAQRSVPAPFLSKTYQLVDDHSTDDVVSWNEEGTAFVWKTAEFAKDLLPQYFKH
NNFSSFIQRLNTYGRKTVDPDKWEFANDYFRGGEDLLDIRRRKSVIASTAGKCVVGS
PSESNSGGDDHGSSTSSPGSSKNPGSVENMVADLSGENEKLKRENNNLSSSELAAAKKQ
RDELVTFLTGHKLVRPEQIDKMIKGGKFKPVESDEESECEGCDGGGGAEEGVGEGCLKLFG
VWLKGERKKRDRDEKNYVVSGRMTEIKNVDFHAPLWKSSKVCN*

>G308 (196..1794)

AGTAATTTAGTTTTTTTTTTTTTTTACAAATTTATTTTGTATTAGAAAGTGGTAGTGG
AGTGAAAAAACAAATCCCTAAGCAGTCCCTAACCGATCCCCGAAGCTAAAGATTCTTCACCT
TCCCAAATAAAGCAAAACCTAGATCCGACATTGAAGGAAAAACCTTTTAGATCCATCTCT
GAAAAAAACCCCAACCATGAAGAGAGATCATCATCATCATCAAGATAAGAAGACTATG
ATGATGAATGAAGAAGACGACGGTAACGGCATGGATGAGCTTCTAGCTGTTCTTGGTTAC
AAGGTTAGGTCATCGGAAATGGCTGATGTTGCTCAGAAACTCGAGCAGCTTGAAGTTATG
ATGTCTAATGTTCAAGAAGACGATCTTTCTCAACTCGCTACTGAGACTGTTCACTATAAT
CCGGCGGAGCTTTACACGTGGCTTGATTCTATGCTACCGACCTTAATCCTCCGTCGTCT
AACGCCGAGTACGATCTTAAAGCTATTTCCCGGTGACGCGATTCTCAATCAGTTCGCTATC
GATTCGGCTTCTTCTCTTAACCAAGGCGGCGAGGAGATACGTATACTACAAACAAGCGG
TTGAAATGCTCAAACGCGCTCGTGGAACACCACAGCGACGCGCTGAGTCAACTCGGCAT
GTTGTCCTGGTTGACTCGCAGGAGAACGGTGTGCGTCTCGTTCACGCGCTTTTGGCTTGC
GCTGAAGCTGTTTCAAGGAGAATCTGACTGTGGCGGAAGCTCTGGTGAAGCAAATCGGA
TTCTTAGCTGTTTTCTCAAATCGGAGCTATGAGACAAGTCGCTACTTACTTCGCCGAAGCT
CTCGCGCGGCGGATTACCGTCTCTCTCCGTCGCAGAGTCCAATCGACCACTCTCTCTCC
GATACTCTTCAGATGCACCTTCTACGAGACTTGTCTTATCTCAAGTTCGCTCACTTCACG
GCGAATCAAGCGATTCTCGAAGCTTTTCAAGGGAAGAAAAGAGTTCATGTCAATTGATTTT
TCTATGAGTCAAGGTCTTCAATGGCCGGCGCTTATGCAGGCTCTTGGCTTCGACCTGGT
GGTCTCTCTGTTTTCCGGTTAACCAGGAAATGGTCCACCGGCACCGGATAATTTGATTAT
CTTCATGAAGTTGGGTGTAAGCTGGCTCATTTAGCTGAGGCGATTACCGTTGAGTTTGGT
TACAGAGGATTTGTGGCTAACACTTTAGCTGATCTTGATGCTTCGATGCTTGAGCTTAGA
CCAAGTGAGATTGAATCTGTTGCGGTTAACTCTGTTTTCGAGCTTCACAAGCTCTTGGGA
CGACCTGGTGCGATCGATAAGGTTCTTGGTGTGGTGAATCAGATTAAACCGGAGATTTT
ACTGTGGTTGAGCAGGAATCGAACATAATAGTCCGATTTTCTTAGATCGGTTTACTGAG
TCGTTGCATTATTACTCGACGTTGTTTGAATCTGTTGGAAGGTGTACCGAGTGGTCAAGAC
AAGGTCATGTCGGAGTTTACTTGGGTAAACAGATCTGCAACGTTGTGGCTTGTGATGGA
CCTGACCGAGTTGAGCGTCATGAAACGTTGAGTCAGTGGAGGAACCGGTTCCGGTCTGCT
GGGTTTGGCGCTGCACATATTGGTTCGAATGCGTTTAAAGCAAGCGAGTATGCTTTTGGCT
CTGTTCAACGGCGGTGAGGGTTATCGGGTGGAGGAGAGTGACGGCTGTCTCATGTTGGGT
TGGCACACACGACCGCTCATAGCCACCTCGGCTTGGAACTCTCCACCAATTAGATGGTG
GCTCAATGAATTGATCTGTTGAACCGGTTATGATGATAGATTTCCGACCGAAGCCAACT
AAATCCTACTGTTTTTCCCTTTGTCACTTGTTAAGATCTTATCTTTCATTATATTAGGTA
ATTGAAAAATTTTAATCTCGCCTAAATTACT

>G308 Amino Acid Sequence (domain in AA coordinates: 270-274)

MKRDHHHHQDKKTMNNEEDDGNGMDELLAVLGKVRSEMAADVAQKLEQLEVMMNSVQ
EDDLSQLATETVHYNPAELYTWLDSMLTDLNPPSSNAEYDLKAIPGDAILNQFAIDSASS
SNQGGGGDTYTTNKRKCSNGVVETTTATAESTRHVVVLVDSQENGVRVLVHALLACAEAVQ
KENLTVAEALVKQIGFLAVSQIGAMRQVATYFAEALARRIYRLSPSQSPIDHSLSDTLQM
HPYETCPYLKFAHFTANQAILFAFGKKRVHVIDFSMSQGLQWPAALMALRPGGPPVF
RLTGIGPPAPDNFDYIHEVGKLAHLAEAIHVEFEYRGFVANTLADLDASMLELRPSEIE
SVAVNSVFELHKLGRPGAIDKVLGVVNQIKPEIFTVVEQESNHNSPIFLDRFTESLHY
STLFDSLEGVPSGQDKVMSEVYLGKQICNVVACDGPDRVERHETLSQWRNRFSGAGFAAA
HIGSNAFKQASMLLALFNGGEGYRVEESDGLMLGWHTRPLIATSAWKLSTN*

>G38 (149..1156)

GAGGAAACTCGAAAAAGCTACACACAAGAAGAAGAAAAGATACGAGCAAGAAGACT
AAACACGAAAGCGATTATCAACTCGAAGGAAGAGACTTTGATTTCAAATTTTCGTCCTCC
TATAGATTGTGTTTCTGGGAAGGAGATGGCAGTTTATGATCAGAGTGGAGATAGAAA

CAGAACACAAATTGATACATCGAGGAAAAGGAAATCTAGAAGTAGAGGTGACGGTACTAC
TGTGGCTGAGAGATTAAAGAGATGGAAAGAGTATAACGAGACCGTAGAAGAAGTTCTAC
CAAGAAGAGGAAAGTACCTGCCAAAGGGTCGAAGAAGGGTTGTATGAAAGGTAAAGGAGG
ACCAGAGAATAGCCGATGTAGTTTCAGAGGAGTTAGGCAAAGGATTTGGGGTAAATGGGT
TGCTGAGATCAGAGAGCCTAATCGAGGTAGCAGGCTTTGGCTTGGTACTTTCCCTACTGC
TCAAGAAGCTGCTTCTGCTTATGATGAGGCTGCTAAAGCTATGTATGGTCCTTTGGCTCG
TCTTAATTTCCCTCGGTCTGATGCGTCTGAGGTTACGAGTACCTCAAGTCAGTCTGAGGT
GTGTACTGTTGAGACTCCTGGTTGTGTTCATGTGAAAACAGAGGATCCAGATTGTGAATC
TAAACCCCTTCTCCGGTGGAGTGGAGCCGATGTATTGTCTGGAGAATGGTGCAGGAGAGAT
GAAGAGAGGTGTTAAAGCGGATAAGCATTGGCTGAGCGAGTTTGAACATAACTATTGGAG
TGATATTCTGAAAGAGAAAGAGAAACAGAAGGAGCAAGGGATTGTAGAAACCTGTCAGCA
ACAACAGCAGGATTTCGTATCTGTTGCAGACTATGGTTGGCCCAATGATGTGGATCAGAG
TCACTTGGATTCTTCAGACATGTTTGATGTCTGATGAGCTTCTACGTGACCTAAATGGCGA
CGATGTGTTTGCAGGCTTAAATCAGGACCGGTACCCGGGGAACAGTGTGGCAACGGTTC
ATACAGGCCCGAGAGTCAACAAAGTGGTTTTGTATCCGCTACAAAGCCTCAACTACGGAAT
ACCTCCGTTTCAGCTCGAGGGAAGGATGGTAATGGATTCTTCGACGACTTGAGTTACTT
GGATCTGGAGAACTAAACAAAACAATATGAAGCTTTTGGATTGATATTGCTTAAATC
CCACAACGACTGTTGATTCTCTATCCGAGTTTGTAGTATATAGAGAACTACAGAACACGT
TTTTTCTTGTATAAAGGTGAAGTGTATATATCGAAACAGTGATATGACAATAGAGAAGA
CAACTATAGTTTGTAGTCTGCTTCTTAAAGTTGTTCTTTAGATATGTTTATGTTTGT
TAACAACAGGAATGAATAATACACACTTGTGAAGCTTTTAAAAAAAAAAAAAAAAAAAA
>G38 Amino Acid Sequence (domain in AA coordinates: 76-143)
MAVYDQSGDRNRTQIDTSRKRKSRSGDGTVAERLKRKWEYNETVEEVSTKKRKVPAKG
SKKGMKMGKGGPENSRCFSFRVQRIRWGWVAEIREPNRGSRLWLGTFTPTAQEAASAYDE
AAKAMYGPLARLNFPRSDASEVTSTSSQSEVCTVETPGCVHVKTEDPDCEKPFSSGGVEP
MYCLENGAEEMKRGVKADKHWLSEFEHNYWSDILKEKEKQEQGIVETCQQQQQDLSVA
DYGWPNDVDQSHLDSSDMFVDVDELLRDLNGDDVFAGLNQDRYPGNSVANGSYRPESQQSG
FDPLQSLNYGIPPFQLEGKDGNGFFDDL SYLDLEN*

>G43 (38..643)

CTCCTGTCTTGTCTAAAGAAAAAGAGAGAGGAAGAAATGGAGACTTTTGAGGAAAGCTC
TGATTTGGATGTTATACAGAAACATCTATTGAAGACTTGATGATCCCTGATGGTTTCAT
TGAAGATTTTGTCTTTGATGATACTGCTTTTGTCTCCGACTCTGGTCTCTAGAACCCTT
TAACCCAGTTCGGAACCTGGAACCTAGTTACCTGTTCTTGATCCAGATTCTATGTCCA
AGAGATTCTGCAAAATGGAAGCAGAATCATCATCATCATCAACAACAACGTACCTGA
GGTTGAGACTGTCTCAAACCGGAAAAAACAAGAGGTTTGAAGAAACGAGACATTACAG
AGGCGTGAGAAGGAGGCCATGGGGGAAATTTGCAGCAGAGATTTCGAGATCCGGCAAAGAA
AGGATCCAGGATTTGGTTAGGCACCTTTTGTAGAGTGATATTGATGCTGCAAGGGCTTACGA
CTATGCAGCTTTTAAAGCTCAGGGGAAGAAAGCTGTTCTCAACTTTCCTTTGGATGCCGG
AAAGTATGATGCTCCGGTCAATTCATGCCGAAAAAGGAGGAGAACCGATGTACCACAGCC
TCAAGGAACAACAAGTACTTCATCATCGTCATCAAACTAATGGGGGAATAGTGATGT
TTAATTAGTATATATAGGTTAATATCTTAAGTATGTGAAGCATCATGTATAGAGCCAAGA
ACCTGTTAGACTAGTGTACTGAAAAGAACTCTTGCAAAATATGTACTAAAGAGTTCCTGT
AACAAATGGAACCTTCTGCGTTTCTCTTGTCTTAAAGAGCTTAAAGTTCTAGAAACAAAGT
TCTTGTCTTTTCCGTTTAAA
AAAAAAAAA

>G43 Amino Acid Sequence (domain in AA coordinates: 104-172)
METFEESDLDVIQKHLFEDLMIPDGFIEDFVDDTAFVSGLWSLEPFNPVPKLEPSSPV
LDPDSYVQEIQLMEAESSSSSSTTSPEVETVSNRKKTKRFEETRHYRGVRRRPWGKFAA
EIRDPAKGSRIWLGTFFESDIDAARAYDYAAFKLGRKAVLNFLDAGKYDAPVNSCRKR
RRTDVPQPQGTSTSSSSSN*

>G536 (1..768)

ATGTCGACAAGGGAAGAGAATGTTTACATGGCGAAATTAGCCGAACAAGCTGAACGTTAC
GAAGAAATGGTTGAATTCATGGAGAAAGTTGCGAAAACCTGTTGATGTTGAGGAACCTTCA
GTTGAAGAGAGGAATCTTCTCTCTGTTGCTTACAAGAACGTGATTGGAGCGAGAAGAGCT
TCGTGGAGAATCATTTCTTCGATTGAGCAGAAAGAAGAGAGCAAAGGGAACGAAGATCAT
GTTGCTATTATCAAGGATTACAGAGGAGAGATTGAATCCGAGCTTAGCAAAATCTGTGAT
GGGATTTTGAATGTTCTTGAAGCTCATCTTATTCCTTCTGCTTCACCAGCTGAATCTAAA

GTGTTTTATCTTAAGATGAAGGGTGATTATCATAGGTATCTTGCTGAGTTTAAGGCTGGT
GCTGAAAGGAAAGAAGCTGCTGAAAGCACTTTGGTTGCTTACAAGTCTGCTTCCGACATT
GCCACTGCTGAGTTAGCTCCTACTCACCCGATAAGGCTTGGTCTTGCACTCAACTTCTCT
GTGTTTTACTATGAAATCCTCAACTCGCTGATCGTGCTTGCAAGCTCGCAAAGCAGGCG
TTTGATGATGCAATCGCTGAGTTAGATACATTGGGTGAGGAATCATACAAGGACAGTACA
CTGATTATGCAGCTTCTTAGAGACAATCTCACTCTCTGGACTTCAGATATGACTGACGAA
GCAGGAGATGAGATTAAAGGAGGCATCAAAGCCCGATGGTGCCGAGTAA

>G536 Amino Acid Sequence (domain in AA coordinates:226-233)
MSTREENVYMAKLAEQAERYEEMVEFMEKVAKTV DVEELSVEERNLLSVAYKNVIGARRA
SWRIISSIEQKEESKGNEDHVAIIKDYRGEIESELSKICDGLNLVLEAHLIPSASPESK
VFYLMKMGDYHRYLAEFKAGAERKEAAESTLVAYKSASDIATAELAPTHPIRLGLALNFS
VFYYEILNSPDRACSLAKQAFDDAIAELDTLGEESYK DSTLIMQLLRDNLTLWTSMDTDE
AGDEIKEASKPDGAE*

>G567 (38..1273)
AAAAAGAAGAAATCAGAAAGTGA AAAAGAGAGCGAGCGATGAACAGTATCTTCTCCATTGA
CGATTTCTCCGATCCTTTCTGGGAAACTCCTCCGATTCTCTCAATCCCGACTCTTCTAA
GCCTGTTACGGCGGATGAAGTTAGCCAGAGTCAACCGGAATGGACTTTCGAGATGTTTCT
CGAAGAGATTTCTTCGTGCGGCGTGAGCTCTGAGCCACTTGTTAAACAACAACGCGAT
CGTCGGTGTTTCTTCGGCGCAATCTCTTCCTTCTGTTTCCGGACAGAATGATTTGAGGA
TGATAGTCGATTTCTGTCGTCGATTCGGGAAATTTGGATTGTGCTGCTCCCATGACGAC
GAAGACGGTGAATGTTGATTCCGATGATTATCGTCGTGTTCTTAAGAACAAGCTTGAGGC
TGAGTGCGCGACTGGTGTTTCTCTTCGGGTGGGTCTGTGAAGCCTGAAGATTGACTAG
TTCTCCAGAACTCAACTTCAACAGTTCAATCCAGTCCCTTACTCAAGGAGAAGTTGG
TGTTACTTCTTCTTACCAGCTGAGGTGAAAAAACTGGTGTATCAATGAAGCAGGTTAC
TAGTGGATCGTCGAGAGAAATTTCTGATGACGAGGACCTTGATGAAGAGAATGAACCAC
CGGTTCTTGAAGCCAGAGGACGTTAAAAAATCTAGAAGGATGCTGTCAAATCGTGAGTC
AGCTAGGCGATCTAGAAGGAGAAAGCAGGAGCAACAAGTGACCTCGAAACACAGGTTAA
TGATCTAAAAGGTGAGCATTATCACTTCTTAAACAAGTGAAGCAATGAATCACAAGTA
TGACGAGGCTGCTGTTGGCAATAGAACTAAAGGCTGACATTGAGACATTAAGAGCTAA
GGTGAAAAATGGCGGAAGAAACCGTGAAGAGAGTAACAGGAATGAATCCGATGCTTCTCGG
AAGATCAAGTGGACATAACAACAACAAGAATGCCAATAACTGGTAACAACAGGATGGA
TTCTTCTAGCATTATTCCAGCTTATCAACCACACTCAAACCTAAACCATATGTCAAACCA
AAACATCGGGATCCCAACCATTTACCTCCAAGACTCGGAAACAATTTGCTGCTCCTCC
ATCCCAAACCAAGCTTCCCTTGAGAGAAATAGAAATGGGCAAAATCACCATGTTACTCC
AAGCGCCAACCCGTATGGCTGGAATACCGAACCTCAGAACGATTCAGCATGGCCGAAAAA
ATGCGTGGACTGATCAAACAAGAAGCGGGTTTCGCACTATATTAATGTCTATGCATCTGT
AATTTGTAAGTGTTATTAAGTTACGAATCATGAGAAAACATCTTGTGAAAATACAGTCTC
ATGGCTTATATATATATATAAGCTCTGTCTTATAACATTACAAGATTCTTATTTGAGAA
CGTCTTTCTATTTATAGCTAATAAAAAAAAAAAAAAAAAA

>G567 Amino Acid Sequence (domain in AA coordinates 210-270)
MNSIFSIDFSDPFWETPPIPLNPDSSKPVTADEVSSQSQPEWTFEMFLEEISSAVSSEP
LGNNNNNAIVGVSSAQSLPSVSGQND FEDDSRFRDRDSGNLDCAAPMTTKTVNVDSDDYRR
VLKNKLEAEACATGVSLRVGSVKPEDSTSSPETQLQPVQSSPLTQGELGVTSSLPAEVKKT
GVSMKQVTS GSSREYSDDEDLDEENETGSLKPEDVKKSRRMLSNRESARRSRRRKQEQ
SDLETQVNDLKGEHSSLLKQLSNMNHKYDEAAVGNRI LKAD IETLRKVMAEETVKRVT
GMNPM L LGRSSGHNNNNRMPITGNRMDSSSIIPAYQPHSNLNHMSNQNIPTILPPRL
GNNFAAPPSTSSPLQIRINGQNHVTPSANPYGWNTPEQNDSAWPKKCV D*

>G680 (338..2275)
CAGTTATCTTCTTCTTCTCTCTGTTTTTAAATTTATTTT TAGAGAATTTTTTTTG
TTTTGCTTCCGATTTGATTATTTCCGGGAACGATGACTTCTCCGGGGAGTTCCCGGTGAG
ATGATAAGTCAGATTGATATTTCTCTCCTCCATGGCTACTCTCAAGGGTTTTGGCTGCG
GTGGATTGCTTTGGTTTTCTCTAGAATCTAAAGAGGTTATCACAACGGCTTTGCAATTTGA
AACTTTT CATGTTTGGGGAGATCAAAGATGGTTTCTTTTTTATACTTTACTTGT TAGAGA
GGATTGGAAGCAGCGAATAGCTGCAACCGGTCTGTTATGGATACTAATACATCTGGAGA
AGAAATTATTAGCTAAGGCAAGAAAGCCATATACAATAACAAAGCAGCGAGAGCGATGGAC
TGAGGATGAGCATGAGAGGTTTCTAGAAGCCTTGAGGCTTTATGGAAGAGCTTGGCAACG
AATTGAAGACATATTGGGACAAAGACTGCTGTT CAGATCAGAAGTCATGCACAAAAGTT

CTTCACAAAGTTGGAGAAAAGAGGCTGAAGTTAAAGGCATCCCTGTTTGCCAAGCTTTGGA
CATAGAAATTCGCCCTCCTCGTCTAAACGAAAACCACTCCTTATCCTCGAAAACC
TGGGAACAACGGTACATCTTCTCTCAAGTATCATCAGCAAAAGATGCAAAACTTGTTC
ATCGGCTCTTCTTTCACAGTTGAATCAGGCGTTCTTGGATTGGAAGAAATGCGGTTCTC
TGAGAAAACATCAACTGGAAGAAAATCAAGATGAGAATTGCTCGGGTGTTCCTACTGT
GAACAAGTATCCCTTACCAACGAAACAGGTAAGTGGCGACATTGAAACAAGTAAGACCTC
AACTGTGGACAACGCGGTTCAAGATGTTCCCAAGAAGACAAAGACAAAGATGGTAACGA
TGGTACTACTGTGCACAGCATGCAAACTACCTTGGCATTTCACGCAGATATTGTGAA
CGGAATATAGCAAAATGCCCTCAAAATCATCCCTCAGGTATGGTATCTCAAGACTTCAT
GTTTCATCCTATGAGAGAAGAACTCACGGGCACGCAAACTTCAAGCTACAACAGCATC
TGCTACTACTACAGCTTCTCATCAAGCGTTTCCAGCTTGTCAATTACAGGATGATTACCG
TTCGTTTCTCCAGATATCATCTACTTCTCCAATCTTATTATGTCAACTCTCTACAGAA
TCTGCGAGCTCATGCTGCAGCTACATTGCTGCTTCCGCTTGGCCTTATGCGAGTGTGCG
GAATCTGGTGATTCAACCCCCAATGAGCTCTTCTCTCCAGTATAACTGCCATTGCG
CGCTGCTACAGTAGCTGCTGCAACTGCTTGGTGGGCTTCTCATGGACTTCTTCTGTATG
CGTCCAGCTCCAATAACATGTGTTCCATTCTCAACTGTTGCAGTTCCAACTCCAGCAAT
GACTGAAATGGATACCGTTGAAAATACTCAACCGTTTGAGAAACAAACACAGCTCTGCA
AGATCAAACCTTGGCTTCGAAATCTCCAGCTTCATCATCTGATGATTAGATGAGACTGG
AGTAACCAAGCTAAATGCCGACTCAAAACCAATGATGATAAAATTGAGGAGGTTGTTGT
TACTGCCGCTGTGCATGACTCAAACTGCCCAGAGAAAATCTTGTGGACCGCTCATC
GTGTGGCTCAAATACACCTTCAGGGAGTGACGCAGAACTGATGCATTAGATAAAATGGA
GAAAGATAAAGAGGATGTGAAGGAGACAGATGAGAATCAGCCAGATGTTATTGAGTTAA
TAACCGTAAGATTAAATGAGAGACAACAACAGCAACAACATGCAACTACTGATTGCTG
GAAGGAAGTCTCCGAAGAGGGTCGTATAGCGTTTTCAGGCTCTCTTTCGCAAGAGAAAGATT
GCCTCAAAGCTTTTCGCTCTCAAGTGGCAGAGAAATGTGAATAGAAAACAAAGTGACAC
GTCAATGCCATTGGCTCCTAATTTCAAAGCCAGGATTCTTGTGCTGCAGACCAAGAAGG
AGTAGTAATGATCGGTGTTGGAACATGCAAGAGTCTTAAACGAGACAGACAGGATTTAA
GCCATACAGAGATGTTCAATGGAAGTGAAAGAGAGCCAAAGTTGGGAACATAAACAATCA
AAGTGATGAAAAGTCTGCAAAAGGCTTCGATTGGAAGGAGAAGCTTCTACATGACAGAC
TTGGAGGTAAAAAATACATCCACATTTTATCAATATCTTTAAATCTAGTGTAGTAG
TTTGCTTCTCCAATCTTTATGAAAGAGACTTTTAAATTTCTTCCGAACATTTCTTTGGT
CATGTCAGGTTCTGTACCATATTACCCCATGTCTTGTCTCTGTCTCTGTTGTGTATGC
TACTTGTGGTCTATATGTCATCTGCTACTACTGTTAATTAACCATTAAGCAATGGATTG
TCTTTA

>G680 Amino Acid Sequence (domain in AA coordinates: 24-70)
MDTNTSGEELLAKARKPYTITKQRRWTEDEHERFLEALRLYGRAWQRIEEHIGTKTAVQ
IRSHAQKFFTKLEKEAEVKGIPVCQALDIEIPPRPKRKPNTPYPRKPGNNGTSSSQVSS
AKDAKLVSASSQLNQAFDLLEKMPFSEKSTGKENQDENCSTVNVKYLPTKQVSG
DIETSKTSTVDNAVQDVPKKNKDKDNDGTTVHSMQNYPWPHFADIVNGNIAKCPQNHP
GMVSDQFMFHPMREETHGHANLQATTASATTTASHQAPPACHSQDDYRSFLQISSTFSL
IMSTLLQNPAHAAATFAASVWPYASVGNSSGDSSTPMSSPPSITAIAAATVAAATAWWA
SHGLLPVCAPAPITCVFSTVAVPTPAMTEMDTVENTQPFQKQNTALQDQTLASKSPASS
SDDSDETGVTKLNADSKTNDKIEEVVTAHVDSNTAQKKNLVDRSSCGSNTPSGSDAE
TDALDKMEKDVEDKETDENQPDVIELNNRKIKMRDNNNNNATTDWKEVSEEGRIAFQ
ALFARERLPQSFSPQVAENVNRKQSDTSMPLAPNFKSQSDCAADQEGVVMIGVTCKSL
KTRQTGFKPKYKRCMEVKESQVGNINNQSDEKVKRLRLEGEAST*

>G867 (64..1098)

CACAACACAAACACATTTCTGTTTTCTCCATTGTTTCAAACCATAAAAAAACACAGAT
TAAATGGAATCGAGTAGCGTTGATGAGAGTACTACAAGTACAGGTTCCATCTGTGAAACC
CCGGCGATAACTCCGGCGAAAAAGTCGTCGGTAGGTAACCTATACAGGATGGGAAGCGGA
TCAGCGTTGTGTTAGATTAGAGAACGGCGTAGAAGCTGAATCTAGGAAGCTTCCGTCG
TCAAAATACAAAGGTGTGGTGCCACAACCAACGGAAGATGGGGAGCTCAGATTTACGAG
AAACACCAGCGGTGTGGCTCGGGACATTCAACGAAGAAGACGAAGCCGCTCGTGCCTAC
GACGTCGCGGTTACAGGTTCCGTCGCGGTGACGCCGTCAAAATTTCAAAGACGTGAAG
ATGGACGAAGACGAGGTCGATTTCTGAATTCTCATTGAAATCTGAGATCGTTGATATG
TTGAGGAAACATACTTATAACGAAGAGTTAGAGCAGAGTAAACGGCGTCGTAATGGTAAC
GGAAACATGACTAGGACGTTGTTAACGTCGGGGTTGAGTAATGATGGTGTCTTACGACG

GGGTTTAGATCGGCGGAGGCACTGTTTGAGAAAGCGGTAACGCCAAGCGACGTTGGGAAG
CTAAACCGTTTGGTTATACCGAAACATCACGCAGAGAAACATTTCCGTTACCGTCAAGT
AACGTTTCCGTGAAAGAGTGTTGTTGAACTTTGAGGACGTTAACGGGAAAGTGTGGAGG
TTCCGTTACTCGTATTGGAACAGTAGTCAGAGTTATGTTTGGACTAAAGGTTGGAGCAGG
TTCGTTAAGGAGAAGAATCTACGTGCTGGTGACGTGGTTAGTTTCAGTAGATCTAACGGT
CAGGATCAACAGTTGTACATTGGGTGGAAGTCGAGATCCGGGTCAGATTTAGATGCGGGT
CGGGTTTTGAGATTGTTTCGGAGTTAACATTTACCGGAGAGTTCAAGAAACGACGTCGTA
GGAAACAAAAGAGTGAACGATACTGAGATGTTATCGTTGGTGTGTAGCAAGAAGCAACGC
ATCTTTCACGCCTCGTAACAACCTCTTCTCTTTTTTTTTCTTTTGTGTTTAAATAATTT
TTAAAAACTCCATTTTCGTTTCTTTATTTGCATCGGTTTCTTTCTTCTTGTTTACCAA
GGTTCATGAGTTGTTTTTGTGTATTGATGAAGTGTAAATTTTATTATAGGATAAATTT
TAAAAA

>G867 Amino Acid Sequence (domain in AA coordinates: 59-124)

MESSSVDESTTSTGSICETPAITPAKKSSVGNLYRMGSGSSVVLDSENGVEAESRKLPS
KYKGVVPQPNGRWGAQIYEKHQRVWLGTFNEDDEAARAYDVAVHRFRRRDAVTNFKDKVM
DEDEVDFLNSHSEIVDMLRKHTYNEELEQSKRRRNGNGNMTRTLTSLGSLNDGVSTTG
FRSAEALFEKAVTPSDVGLNRLVLPKHAEKHFPLPSSNVSVKGVLLNFEDVNGKVWRF
RYSYWNSSQSYVLTGKWSRFVEKNLRAGDVVSFSSRNGQDQQLYIGWKSRSGLDLADR
VLRLFGVNIPESSRNDVVGKRVNDTEMLSLVCSKKQRIHAS*

>G956 (1..840)

ATGGAGGAGACAGAAAAGAATAAGGCGCAGCATAAGTATGGTTGAGGCTAATCTACCTCCT
GGTTTTAGATTCCATCCTAGAGACGACGAGCTCGTCTGTGACTACTTAATGAGAAGAACC
GTTTCGACGCTCTATCAACAGTTGTCTTGATCGACGTCGATCTTAACAAATGCGAGCCT
TGGGACATTCTCAACCGCGAGAGTGGGAGGAAAGAATGGTACTTTTACAGCCAAAAA
GACCGCTAAATACGCAACAGGCTACAGAACAAACCGGGCTACGGCCACCGGTTATTGGAAA
GCCACCGGGAAGATAGAGCAATCCAAAGAAACGGTGGTCTTGTGGGTATGAGAAAGACA
CTTGTGTTTTTACCGAGGTCGATCCCTAAAGGTCGTAAACTGATTGGGTCATGCATGAG
TTTCGTCTCCAAGAAAACCTTCTTCACTCCCTAATTCTCTCGAGGAAGAGTGGGTA
TTGTGTAGAGTTTTCCACAAGAACAGCAACCGAGCTGATATAGACGACATCACAAGGAGC
TGCTCTGATGCAACAGCTTCTGCATTCTGACTCTTACATCAACTTCGACCATCATCAC
ATCATCAATCAGCATGTACCCTGCTTCTCCAATAATTTGTCACATAACCAACCAACCAA
TCCGGTTTAATCTCCAAGAACTCCAGCCATTGTTTAAATGCTTCCCTGATCAAATGATT
CTCAGAACTTTGCTAAGTCAACTCACAAAAAAGTCGAAGAATCACAGAGTCGTGGAGAC
GGAAGCTCAGAGAGCCAATTGACCGACATTGGCATCCCAAGCCATGCATGGAATTACTGA

>G956 Amino Acid Sequence (domain in AA coordinates: TBD)

MEETKKNKSISMVEANLPPGFRFHRDDELVCYLMRRTVRSLYQPVVLDVLDLNKCEP
WDIPQTARVGGKEWYFYSQKDRKYATGYRTNRTATGYWKATGKDRAIQRNGGLVGMKRT
LVFYRGRSPKGRKTDWMHEFRLQGKLLHHSNPLEEELVLCRVFHKNSNGADIDITRS
CSDATASAFMDSYINFDDHHIINQHVPCFSNNLSHNQTNQSLISKNSSPLFNASPDQMI
LRTLQLTKKVEESQSRGDSSESQSLTDIGIPSHAWNY*

>G996 (53..1063)

CGATCGATCTTGAATTGATTCTTTGTAGTATTTTATTACATATATATATAGATGGGAAG
ACATTCATGTTGTTACAAACAGAACTGAGGAAAGGACTTTGGTCTCCTGAAGAAGATGA
GAAGCTTCTTTCGTTACATCACTAAGTATGGTCATGGTTGCTGGAGCTCTGTCCCTAAACA
AGCTGGTTTACAGAGATGTGAAAAAGTTGTAGATTAAAGATGGATAAATTATTAAAGACC
AGATTTGAAGAGAGGAGCATTTTCTCAAGATGAAGAAATCTCATTATTGAACCTTCATGC
CGTTCCTGGCAATAGATGGTCTCAGATAGCTGCACAGCTTCTGGAAGAACCACAAATGA
AATCAAGAATCTTTGGAATCTTGTGTTGAAGAAGAAATTGAGGCTGAGAGGAATTGACCC
GGTTACACACAAGCTCTTAACCGAAATCGAAACCGGTACAGATGACAAAACAAAACCGGT
TGAGAAGAGTCAACAGACTTACCCTCGTTGAGACTGATGGCTCCTCTAGTACCACTACTTG
TAGTACTAACCACAAAACAACTGATCATCTTTATACCGGAAATTTCCGGTTTTCAACG
GTTAAGTCTAGAAAACGGTTCAAGAATCGCAGCCGGTTCTGACCTCGGTATCTGGATTCC
CCAAACCGGAAGAAACCATCATCATCATGTGATGAAACCATCCCTAGTGCAGTGGTACT
ACCCGGTTCAATGTTCTCATCCGGTTTAACCGGTTATAGATCCTCCAATCTCGGTTTAAAT
TGAATTGGAAAACCTATTCTCAACCGGGCCAATGATGACAGAGCATCAGCAAATTCAGA
GAGTAACCTACAACAATTCAACATTTCTTGGAAATGGGAATCTGAATTGGGGATTACAAT
GGAGGAAAATCAAAATCCATTACAAATATCGAATCATTCAAATTCGTCTTATACAGTGA

TATAAAATCAGAGACCAATTTTTTTGGCACAGAGGCTACAAATGTTGGTATGTGGCCATG
TAACCAGCTTCAGCCTCAGCAACATGCATATGGCCATATATAAATCTTCTGTATATTAT
AA

>G996 Amino Acid Sequence (domain in AA coordinates: 14-114)
MGRHSCCYKQKLRKGLWSPEEDEKLLRYITKYGHGCWSSVPKQAGLQRCGKSCRLRWINY
LRPDLKRGAFSQDEENLIIELHAVLGNRWSQIAAQLPGRTDNEIKNLWNSCLKKKLRRLRG
IDPVTHKLLTEIETGTDKTKPVEKSQQTLYLVETDSSSTTTCSTNQNNNDHLYTGNFG
FQRLSLENGSRIAAGSDLGWIWIPQGRNHHHHVDETIPSAVVLPGSMFSSGLTGYRSSNL
GLIELENSFSTGPMTEHQIQESNYNNSTFFNGNLNLWGLTMEENQNPFITSNHSNSSL
YSDIKSETNFFGTEATNVGMWPCNQLQPQGHAYGHI*

>G1946 (90..1547)
TCTCACCTATTGTAAAAATCACCAGTTTCGTATATAAAACCCCTAATTTTCTCAAAATTC
CAAATATTGACTTGAATCAAAAATCCGAATGGATGTGAGCAAAGTAACCACAAGCGACG
GCGGAGGAGATTCAATGGAGACTAAGCCATCTCCTCAACCTCAGCCTGCGGCGATTCTAA
GTTCAAACGCGCCTCCTCCGTTTCTGAGCAAGACCTATGATATGGTTGATGATCACAAATA
CAGATTGATTTGCTCTTGGAGTGCTAATAACAACAGTTTTATCGTTTGGAAACCAACCGG
AGTTCGCTCGCGATCTTCTTCCTAAGAAGCTTAAAGCATAATAATTTCTCCAGCTTCGTTA
GACAGCTTAATACCTATGGTTTCAGGAAGGTTGACCCAGATAGATGGGAATTTGCGAATG
AAGGTTTTTAAAGAGGTCAGAAGCACTTGCTACAATCAATACTAGGCGAAAACCTGCCC
ATGGACAGGGACAGGGACATCAGCGATCTCAGCACTCGAATGGACAGAACTCATCTGTTA
GCGCATGTGTTGAAGTTGGCAAATTTGGTCTCGAAGAAGAAGTTGAAAGGCTTAAAGAG
ATAAGAACGTCCTTATGCAAGAAGCTCGTCAGATTAAGACAGCAGCAACAGTCCACTGATA
ACCAACTTCAAACGATGGTTTCAGCGTCTCCAGGGCATGGAGAATCGGCAACAACAATTAA
TGTCATTCTTGCAAAGGCGAGTACAAAGCCCTCATTTTCTATCTCAATTCTTACAGCAGC
AGAATCAGCAAAACGAGAGTAATAGGCGCATCAGTGATACCAGTAAGAAGCGGAGATTCA
AGCGAGACGGCATTGTCCGTAATAATGATTCTGCTACTCCTGATGGACAGATAGTGAAGT
ATCAACCTCCAATGCACGAGCAAGCCAAAGCAATGTTTAAACAGCTTATGAAGATGGAAC
CTTACAAAACCGGCGATGATGGTTTCTTCTAGGTAATGGTACGTCTACTACCGAGGGAA
CAGAGATGGAGACTTCATCAAACCAAGTATCGGGTATAACTCTTAAGGAAATGCCTACAG
CTTCTGAGATACAGTCATCATCAACCAATTGAAACAACCTCTGAAAATGTTTCGGCAGCAT
CAGAAAGCAACCGAGAAGTGTATTCCTTACCTGATGATCTAACTCTTCCCGACTTCACTC
ATATGCTACCGGAAAATAATTCAGAGAAGCCTCCAGAGAGTTTCATGGAAACCAACCTGG
GAGGTTCTAGTCCATTACTAGATCCAGATCTGTTGATCGATGATTCTTTGTCCTTCGACA
TTGACGACTTTCCAATGGATTCTGATATAGACCCTGTTGATTACGGTTTACTCGAACGCT
TACTCATGTCAAGCCCGGTTCCAGATAATATGGATTCAACACCAGTGGACAATGAAACAG
AGCAGGAACAAAATGGATGGGACAAAACCTAAGCATATGGATAATCTGACTCAACAGATGG
GTCTCTCTCTCTGAAACCTTAGATCTCTCAAGGCAAAATCCTTGATTTTGGGAGTTTT
TAAAGCTTTTTGAGGTAACACAGTCCCTGAGAGCAGCATATTCAT

>G1946 Amino Acid Sequence (domain in AA coordinates: 32-130)
MDVSKVTTSDGGGDSMETKPSQPQPAAILSSNAPPPFLSKTYDMVDDHNTDSIVSWSAN
NNSFIVWKPEFARDLLPKNFKHNMFSSFVRQLNTYGFRRKVDPRWEFANEGFLRGQKHL
LQSITRRKPAHQGGQGHQRSQHSNGQNSSVSACVEVGKFGLEEEVERLKRDKNVLQELV
RLRQQQSTDNQLQTMVQRLQGMENRQQQLMSFLAKAVQSPHFLSQFLQQNQNESNRR
ISDTSKKRRFRKRDGIVRNDSATPDGQIVKYQPPMHEQAKAMFKQLMKMEPYKTGDDGFL
LGNGTSTTEGTEMETSSNQVSGITLKEMPTASEIQSSSPIETTPENVSAASEATENCIPS
PDDLTLPDFTHMLPENNSEKPPESFMEPNLGGSSPLLDPLLIDDSLSFDIDDFPMDSDI
DPVDYGLLERLLMSSPVPDNDSTPVDNETEQEQNGWDKTKHMDNLTQQMGLLSPETLDL
SRQNP*

>G217 (84..2618)
cttcgttcttaccgaggtccacgagcattagcttcagagacctgaattggagtgcggtt
ggatcaaaaacagttgagcgaagatgaggattatgattaagggaggtggttggagaaca
ccgaagatgagattctcaagccgctgatgaagtatggtaagaaccaatgggctcggg
tctcgctcttctcggttcgtaagtctgctaaacagtgtaagctcgctggtacgagtggt
tcgatccatctatcaaaaagactgaatggaccagagaagaagatgagaagcttctacatc
ttgctaaacttctgcctactcaatggagaactattgctcctattgtgggtcgtacaccat
ctcaatgtccttgagaggtatgagaagctccttgatgcagcatgcactaaggatgaaaatt
atgatgcagcggtatgatccacgaaaattacgtcctggtgagattgatccgaaccagaag

caaagcctgctcgtcctgatccggtagacatggacgaagatgagaaagaaatgctttctg
aagcaagagctagatttggttaacacgaggggaaagaaggctaaaagaaaagctagagaaa
aacaacttgaggaagctagaaggcttgcttctctgcaaaaaagaagagaactaaaagcag
ctgggattgatggaaggcatagggaaagaaagagaaagggaatcgactataatgcagaaa
ttccttttgaaaagaggcaccctgcggttcttatgatactgaggatgaagatcgctctg
ctgatcaagtaaaatttccaactaccattgaagaacttgaaggaaaaagaagagctgatg
tagaagcacatttacgcaacaagatgttgcaaggaaataaaattgctcagagacaggatg
ctccagcagctatattgcaagcaacaagctgaatgatccggaagttgttaggaagaggt
caaagctgatgttaccaccaccgcagatttccagaccagagctagaagaaattgctaaga
tggtgatgccagtgaccttcttgccgagaatgaggagctaacagaaggcagtgctgcta
ctcgtgaccttttgcaaaattactcacaacaccaaggcaaggaaatgacacccatgagga
cacctcaaagaactcctgctggttaaaggtgatgctattatgatggaagcagaaaacctgg
ccagattaagagactctcagacacctttgctaggaggagaaaaatcctgagttgcaccctt
ctgacttctactggggtcactccgagaaagaaggagattcaaacgcctaatacaatgttga
cccttcaatgactcctggtggtgctggtcttactccaagaattggcttgacgccatcaa
ggatgggtccttcttttctatgacacccaaaggactcccttcagggatgaacttcaca
ttaacgaagacatggacatgcagcaaaagtgcaaaacttgagaggcagagacgagaggaag
ctagaaggagtttacgctctggtttgactgggttctcctcagccaaagaacgagtaccaa
tagttgcacaacctcctcctgaggaaagtgaagagccagaagagaaaaattgaggaagaca
tgtcagacaggatagcgagggaaaggcgagggaagaagcaagacaacaggcattgctta
agaagagatccaaggtcctgagagagatcttctcagacccccagctgcttcattggcag
taattaggaactcggtgctttcagctgatggagacaaaagttctgttctcctcactc
cgattgaggttgagataaaaatggtaagagaggagcttctacagttgctggagcatgata
atgcaaagtatccgcttgatgacaaagctgagaagaagaaggagccaagaaccgtacca
accgttctgcttctcaagttctgcaattgacgattttgatgaaaatgagctccaagagg
ctgacaaaaatgataaaggaggaggggaagttctgtgtgtgtaatgggacatgagaaca
agacacttgatgattttgtagaagctcacaacacatgcgtgaatgatctcatgtatttcc
ccactcgaagcgcttacgagctctcaagtgttgctgggaacgaggacaaagttgcagctt
ttcaggaggagatggagaatgtgagaaaaaagatggaggaggatgagaagaaggcagaac
acatgaaggccaagtacaaaacttatacaaagggctcatgagaggagggcagagaccgtgt
ggacccaaatagaggcgacattgaagcaggctgagattggtggaaagagaagtagagtgt
ttaaagcattgaaggaggcaagaagagatggctgcatcttttaggaaaaagaatttgcaag
aggaagtataaagcaaaaggaaacagagagtaaaactgcagactcgctatgggaatatgt
tggcaatgggtgaaaaagcagaggagataaatggctcggtttccgagcacaggcattgaaga
aacaagaggatgttgaagatttccaaaactgaaagaagctaaagctagccactggagagg
aagaggacatagccatagccatggaagcttctgcataaaaacttgagttttgtattgctt
acaagttttaaggagacgtagcttgactttgtattggttaagtttttttaatatgagtc
gactttgtaaaaaggttatgatataattctctggtttgtatgctttgcaagagtcaga
aat
ttgaatgcttcaggatcaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
>G217 Amino Acid Sequence (conserved domain in AA coordinates: 8-67)
MRIMIKGGVWKNTEDEILKAAVMKYGKNQWARISSLLVRKSAKQCKARWYEWLDPSIKKT
EWTREDEKLLHLAKLLPTQWRTIPIVGRTPSQCLERYEKLLDAACTKDENVDAADDP
KLRPGEIDPNPEAKPARPDPVMDDEDEKEMLSEARARLANTRGKKAKRKAREKQLEEARR
LASLQKRRELKAAGIDGRHRKRKRKGIDYNAEIPFEKRAPAGFYDTADED RPADQVKFPT
TIELEGKR RADVEAHLRKQDVARNKIAQRQDAPAILQANKLNDPEVVRKRSKLM LPPP
QISDHELEELIAKMGYASDLLAENEELTEGSAATRALLANYSQTPRQGMT PMRTPORTPAG
KGDAIMMEAEENLARLRDSQTPLLGENPELHPSDFTGVT PRKKEIQTPNPMLTPSMT PGG
AGLTPRIGLTPSRDSSFSMT PKGTPFRDELHINEDMDMQQSAKLERQRREEARRSLRSG
LTGLPQPKNEYQIVAPPPPEESEPEEKIEEDMSDRIAREKAEERQALLKKRSKVLQ
RDLPRPPAASLAVIRNSLLSADGDKSSVVPPTPIEVADKMVREELLQLLEHDNAKYPLDD
KAEKKGAKNRTNRSASQVLAIDDFDENELQEADKMIKEEGKFLCVSMGHENKTLDDFVE
AHNTCVNDLMPFPTRSAYELSSVAGNADKVAAFQEEMENVRKKMEEDEKKAHEMKAKYKT
YTKGHERRAETVWTQIEATLKQAEIGGTEVECFKALKRQEEMAASFRKKNLQEEVIKQKE
TESKLQTRYGNMLAMVEKAEIIMVG FRAQALKKQEDVEDSHKLKEAKLATGEEEDIAIAM
EASA*
>G2192 (92..2971)
CGGAAAGAGATCAACCAACGATAGAGGAGAAGAAGAACTGCATACGCAAAAAA ACTTTC

CCGGGAAAATTCCAGAACTGCTTTGGAAAAATGTGCGAGCCCGATGATAATTCCGCTAG
AAACGGCGTCACTACTCAACCTTCGAGGTCAAGGGAGCTTCTAATGGATGTTGACGACTT
AGATCTTGACGGTTCATGGCCACTAGATCAAATCCCTTACTTATCCTCATCGAATCGCAT
GATTTCTCCGATTTTTGTCTCCTCTTCTCTGAGCAGCCTTGCTCGCCTCTCTGGGCTTT
CTCCGACGGTGGAGGAAATGGTTTTACCACGCAACCTCCGGTGGCGATGATGAGAAGAT
CAGCTCTGTCTCCGGTGTTCTTCTTCCGTCTCGCCGAGTATCCTCTCTTCTCCCTTA
CTTTCTCCATCAGCAGCTGAGAACACAACAGAGAAGCATAACAGTTTCCAGTTTCCGTC
TCCATTGATGAGCCTAGTCCCACCAGAGAACACAGACAACTACTGTGTGATCAAAGAGAG
GATGACTCAGGCGCTTCGATACTTCAAAGAATCAACCGAACAACACGTTTTGGCTCAGGT
CTGGGCTCCTGTGAGAAAGAATGGTCGTGATTTGCTGACGACTTTGGGTCAACCTTTTGT
TCTTAATCCTAATGGTAATGGGCTTAATCAATACAGGATGATCTCTCTCACATATATGTT
TTCTGTGGATAGTGAAAGTGACGTAGAGCTCGGACTCCCGGGTCGAGTTTTCCGTCAGAA
ATTGCCTGAATGGACTCCAAATGTTTCAGTACTATTCCAGCAAAGAATTCTCGCGGCTTGA
TCACGCCCTTGCACTACAACGTGCGTGGTACACTGGCCTTGCTGTCTTTAATCCCTCTGG
TCAGTCTTGCAATAGGTGTTGTGGAACCTTATAATGACCTCAGAGAAGATTCACTATGCACC
CGAAGTGGACAAAGTTTGCAAAGCCTTGAGGCGGTAAATCTGAAAAGCTCGGAAATACT
TGATCACCAAACAACACAGATATGCAATGAGAGTCGCCAAAACGCGCTTGCTGAGATTCT
CGAAGTGCTGACAGTTGTATGTGAGACCCATAACTTGCCTCTCGCTCAGACTTGGGTTC
ATGTCAGCATGGGAGCGTTCTTGCCAATGGTGGCGGTCTAAAGAAAACTGCACCAGCTT
TGACGGTAGCTGCATGGGTCAAATCTGCATGTCTACAACCGACATGGCCTGCTATGTCGT
GGATGCTCATGTCTGGGGCTTTAGAGATGCCTGTCTTGAACACCATCTCCAGAAAGGCCA
GGGAGTCGCTGGACGAGCTTTTCTCAATGGTGGCTCATGTTTCTGCAGAGACATCACCAA
GTTCTGCAAAACGCAGTACCCACTAGTCCATTATGCGCTCATGTTCAAGTTGACCACTTG
TTTTGCAATATCTCTCCAGAGCTTTACACGGGCGACGACAGTTACATTCTTGAATTTTT
TCTTCCTTCGAGTATAACAGACGCAAGAGCAAGATTGCTGTTGGGTTCTATTTTGGT
GACAATGAAAGAACATTTTCAGAGTCTGAGGGTTGCATCTGGGGTTGACTTTGGTGAAGA
TGACGACAAATGTCTTTTCGAGATCATCCAAGCATTACCGGACAAGAAGGTTTCATTCAA
AATAGAATCCATTTCGAGTTCCCTTTTCTGGTTTTAAGTCAAATGCAACAGAGACGATGTT
GATTCCTCAGCCTGTGGTTCACTTCTGATCCAGTAAATGAGAAAATCAACGTGGCCAC
TGTTAACGGTGTGGTTAAGGAGAAGAAGAAAACAGAGAAAAGCGTGGGAAGACTGAGAA
AACAATCAGTCTAGATGTACTTCAGCAGTATTTCACTGGGAAGTCTCAAAGACGCTGCAAA
GAGCCTAGGAGTTTGCCCGACGACAATGAAGCGAATTTGCAGGCAACACGGAATCTCGCG
GTGGCCATCGAGGAAGATCAAGAAAGTGAATCGTTCAATCACAAGCTGAAACGAGTCAT
CGAATCTGTTCAAGGTACTGATGGAGCCTCGACCTGACTTCCATGGCCGTTAGTTCCAT
CCCTTGGACACACGGTCAAACATCAGCACAGCCACTAACTCACCCAATGGTTCCAAACC
ACCTGAGCTACCAAACACCAATAATTCACCTAACCATTTGGTCAAGTGATCACAGTCCGAA
CGAGCCAAATGGTTTCGCTGAGTTACCACCAAGCAATGGTCAAGCGATCACGAACGGT
GGATGAGAGCGCTGGGACTCCAACCTCTCATGGCTCATGTGACGGTAACCAATTAGATGA
ACCGAAAGTCCCAAATCAAGATCCGCTCTTACGGTTGGTGGATCACCCGGGCTCCTTTT
TCCACCTTATCTAGAGATCATGATGTATCTGCAGCTTCTTCGCAATGCCGAACAGGCT
TCTTGGTTCTATAGACCATTTCCGAGGAATGCTCATTGAAGACGCTGGAAGTTCAAAGA
TCTGAGAAATCTCTGCCCCACTGCAGCATTTGACGATAAGTTTCAAGACACAACTGGAT
GAACAATGATAATAATAGCAACAACTTATACGCTCCCCCAAAGGAAGAGGCCATTGC
AAATGTTGCATGCGAACCATCAGGCTCAGAAATGAGAACGGTAACAATCAAAGCAAGTTA
CAAAGACGACATAATACGGTTTCAAGATATCCTCGGGTTCAGGTATAATGGAATTGAAGGA
TGAAGTGGCTAAGAGGCTGAAAGTTGATGCAGGAACGTTTCGATATCAAGTATCTTGACGA
TGATAACGAATGGGTTTTAATAGCTTGTGATGCTGATCTTCAAGAATGTCTCGAGATCCC
TAGATCCTCCCGCAEGAAAATCGTAAGGCTCTTAGTTTCATGATGTAACGACAAATCTAGG
GAGCTCCTGCGAGAGCACTGGAGAATTGTGACCTGATAATTCATTTCGAACCTCTTTGTAA
ATAG

>G2192 Amino Acid Sequence (conserved domain in AA coordinates:600-700)

MCEPDDNSARNGVTTQPSRSRELLMDVDDLDLDSWPLDQIPYLSSSNRMISPIFVSSSS
EQPCSPLWAFSDGGNGFHHATSGGDDEKISSVSGVPSFRLAEYPLFLPYSSPSAAENTT
EKHNSFQFPSPMLSLVPPENTDNYCVIKERMTQALRYFKESTEQHVLAQVWAPVRKNGRD
LLTTLGQPFVLNPNGLNQLYRMISLTVMFSVDSDELGLPGRVFRQKLPEWTFPNVQY
YSSKEFSRLDHALHYNVRGTLALPVFNPSGQSCIGVVELIMTSEKIHYAPEVDKVCALAE
AVNLKSSEILDHQTQICNESRQNALAEILEVLTVCETHNLPLAQTWVPCQHGSVLANG

GGLKKNCTSFDGSCMGQICMSTTDMACYVVDHAVHWGFRDACLEHHLQKGQGVAGRAFLNG
 GSCFCRDITKFKCTQYPLVHYALMFKLTTCAISLQSSYTGDDSYILEFFLPSSITDDQE
 QDLLLGSILVTMKEHFQSLRVASGVDFGEDDDKLSFEIIQALPDKKVHVKIESIRVPFSG
 FKSNAETETMLIPQPVVQSSDPVNEKINVATVNGVVKEKKTEKKRGKTEKTISLDVLQOY
 FTGSLKDAAKSLGVCPTTMKRICRQHGISRWPSRKIKKVNRSITKLKRVIESVQGTGGGL
 DLTSMAVSSIPTWHTGQTSAPQLNSPNSGKPPPELNTNNSPNHWSSDHSPPNEPNSPELPP
 SNGHKRSRTVDESAGTPTSHGSCDGNQLDEPKVPNQDPLFTVGGSPGLLFPYSRDHDVS
 AASFAMPNRLGSLDHFRGMLIEDAGSSKDLRNLCPATAFDDKFQDTNWMNNDNNSNNNL
 YAPPKEEAIANVACEPSGSEMRTVTIKASYKODIIRFRISSGSGIMELKDEVAKRLKVDA
 GTFDIKYLDLDDNEWVLIACDADLQECLEIPRSSRTKIVRLLVHDVTNGLSSCESTGEL*

>G504 (69..1040)

CGTCGACCTCTTGACGATCATGAGACTGATTTCTGTGAAAATATCGTCATTATATCAAATT
 AGAAGTTGATGGAAAACATGGGGGATTTCGAGCATAGGGCCGGGCCATCCGCATCTCCCTC
 CCGGGTTTCGGTTTTCACCCGACTGATGAGGAAGTAGTAGTTTATTACCTCAAGAAGAAAG
 CAGATTCTGTTCCACTTCCAGTCTCAATCATCGCAGAGATTGATCTTTACAAGTTTGATC
 CTTGGGAGCTTCCAAGCAAGGCGAGTTTGGAGAGCACGAGTGGTACTTCTTTAGTCCTC
 GGGATCGGAAGTATCCAAATGGGGTTAGGCCAAACCGGGCAGCAACTTCCGGTTATTGGA
 AAGCAACGGGAACCGATAAACCGATATTTACGTGCAATAGTCACAAGGTTGGTGTCAAGA
 AAGCGCTTGTTTTTACCGTGGAAAGCCTCCTAAAGGGATAAAAACAGATTGGATCATGC
 ATGAATATCGCCTCACTGATGGTAACCTTAGCACTGCGGCTAAGCCGCTGACTTAACCA
 CGACAAGGAAAACTCACTACGGCTAGACGATTGGGTTCTATGTAGGATCTATAAGAAGA
 ATAGTTCACAAAGACCAACAATGGAGAGAGTATTACTTAGAGAGGATCTAATGGAAGGCA
 TGCTCTCAAAATCATCTGCTAATTCTTCTTCTACATCAGTACTAGACAACAACGACAACA
 ATAATAACAATAACGAAGAACAACCTTTTTCGACGGTATGGTCTGTTTCTTCAGACAAACGTT
 CCTTGTGTGGTCAATACCGAATGGGCCACGAGGCTCAGGATCATCTTCATTCCGATCTT
 TCTTATCGAGCAAGAGGTTTCATCATACAGGTGATCTCAACAATGATAACTACAATGTCT
 CTTTTGTTTCGATGCTTAGTGAGATTCTCAGAGTTCCGGGTTTCATGCAAATGGTGTTA
 TGGATACGACGTCGTCTCTAGCTGATCATGGGGTTTTAAGACAGGCGTTTCAGCTTCCTA
 ACATGAAGTGGCACTCATAATCTATATAGATATATATGTGTGTATCATATATGTATCTAT
 GCAGGCCTAATATAGTTTACACATAAATCATCTGGGGCGGCCGCT

>G504 Amino Acid Sequence (domain in AA coordinates: TBD)

MENMGDSISIGPHPLPPGFRFHPTDEELVVHYLKKKADSVPLPVSIIAEIDLYKFDPE
 LPSKASFGHEHWYFFSPDRKYPNGVRPNRAATSGYWKATGTDKPIFTCNHVKGVKKAL
 VFYGGKPPKGIKTDWIMHEYRLTDGNLSTAAPDLTTTRKNSLRLLDDWVLCRIYKKNSS
 QRPTMERVLLREDLMEGMLSKSSANSSTSVLDNNDNNNNNNEEHFFDGMVVSDDKRSCL
 GQYRMGHEASGSSSFGSFLSSKRFHHTGDLNNDNYNVSVFVSMLEIPQSSGFHANGVMDT
 TSSLADHGVLRQAFQLPNMNWHS*

>G622 (248..2620)

TCTTTCTTTCTCAATTCGCGCTCAAAATCTTCTCTTTCTTTCCCCCGCGGTCTTCA
 CCAATCCTCTGATCTCTACACACGAACCTTTGATTTTGACCAACGTCGATGCATGTTT
 ATGACTAGTCTCTTCTCAATCTTCAATTTTCATCAATTCACGTCGATTTCTGATCCGAT
 TCGTTGTTCTAGCTCTTTGTGTGGTGTAGGGTTTTAAGATTTTGAATTGGGGTTTGA
 GTTTGTGATGTTTGAAGTCAAAATGGGGTCAAAGATGTGCATGAACGCTTCATGTGTAC
 GACTTCTACTGTTGAATGGAAGAAAGGTTGGCCTCTTCGATCTGGTCTTCTCGCTGATCT
 CTGTTATCGTTGCGGATCTGCGTATGAGAGTTCTCTATTCTGTGAACAATTTTATAAGGA
 CCAATCTGGTTGGAGGGAATGCTATTTGTGTAGCAAGAGACTACATTGTGGATGCATTGC
 TTCTAAGGTAACGATTGAGTTAATGGACTATGGTGGTGTGGTTGTAGTACATGTGCTTG
 CTGCCATCAACTCAATTTGAACACAAGGGGTGAGAATCCAGGTGTTTTTAGCAGATTGCC
 AATGAAAACGTTAGCTGATAGGCAACATGTAAATGGCGAAAGCGGAGGAAGAAACGAAGG
 CGATCTCTTTCTCAGCCACTAGTCATGGGCGGAGATAAAGGGAAGAGTTTCATGCCTCA
 CCGTGGGTTTGGTAAGCTAATGAGTCCAGAAAGTACAACCACCGGCATAGGCTGGATGC
 TGCTGGGGAAATGCATGAATCATCACCTTTACAGCCATCTTTAAATATGGGTTTGGCTGT
 GAATCCGTTTAGCCCATCTTTTGAACCGAGGCTGTGAGGGGAATGAAACACATCAGTCC
 TTCTCAGTCCAACATGGTCCATTGCTCTGCTTCTAATATACTGCAAAAGCCATCAAGACC
 TGCTATTTCAACTCCTCTGTGGCTAGTAAATCCGCTCAGGCGCGGATTGGAAGGCCCTCC
 TGTCGAAGGGCGAGGGAGAGGCCACTTGCTTCCGCGGTATTGGCCAAAATATACGGATAA
 AGAGGTTTCAGCAGATCTCTGGAAATTTGAATTTGAACATTGTACCTCTCTTTGAGAAAAC

TCTTAGTGCCAGTGATGCTGGTCGCATTGGTTCGTCTAGTTCTTCCAAAAGCCTGTGCAGA
 GGCATATTTTCTCCGATTAGTCAATCCGAAGGCATTCCTTTGAAAATCCAAGATGTGAG
 GGGTAGGGAGTGGACGTTCCAGTTCAGATATTGGCCCAATAACAATAGTAGAATGTATGT
 TTTAGAAGGTGTCACTCCATGCATACAGTCCATGATGCTACAGGCTGGTGATACAGTAAC
 TTTCACTCGGGTTGATCCTGGCGGAAAACTAATCATGGGTTCAGGAAGGCAGCTAATGC
 TGGAGACATGCAGGGTTGTGGGCTCACCAACGGAACATCAACTGAGGACACATCATCGTC
 TGGTGTAAACAGAAAACCCACCCTCCATAAATGGTTCTCGTGTATTTCACTAATACCGAA
 AGAGTTGAATGGTATGCCTGAGAATTTGAACAGTGAGACTAACGGGGGCGAGGATAGGTGA
 TGATCCTACACGAGTTAAAGAGAAGAAGAGAACTCGAACCATTGGTGCAAAAAATAAGAG
 ACTTCTTTTGCATAGTGAAGAATCTATGGAGCTGAGACTCACTTGGGAAGAAGCTCAGGA
 CTTGCTTCGTCCCTCTCCTAGTGTAAGCCTACCATCGTTGTCTATGAGGAGCAAGAAAT
 TGAAGAATATGACGAACCTCCTGTCTTTGGAAAGAGGACTATAGTCACTACAAAACCTTC
 AGGTGAACAGGAACGATGGGCAACTTGGCAGCAGTCTCTAAATGGAGAAGGTTACCTGT
 AGATGCTCTTCTTTCTTTAAATGGACATGTATAGACAATGTTTGGGATGTGAGTAGGTG
 TTCATGTTCTGCACCGGAGGAGAGTCTGAAGGAACCTGAGAATGTTCTTAAAGTAGGTAG
 AGAGCACAAGAAGAGAAGAACTGGGGAAAGACAGGCAGCACAAAGTCAGCAAGAACCGTG
 TGGTTTGGACGCACTGGCGAGTGCAGCAGTCTTAGGAGACACAATAGGCGAGCCAGAGGT
 AGCGACCACGACCAGACATCCAAGGCACAGGGCTGGATGCTCTTGCATCGTGTGCATTCA
 GCCACCAAGTGGGAAAGGTAGGCACAAGCCTACATGTGGCTGCACTGTGTGTAGCACCGT
 GAAGAGAAGGTTCAAGACGCTTATGATGAGGAGGAAGAAGAAGCAGTTGGAGCGCGATGT
 AACAGCAGCAGAAGATAAGAAGAAGAAGGACATGGAACCTGGCTGAGTCTGATAAGAGTAA
 GGAGGAGAAGGAAGTGAACACAGCGAGAATAGACCTGAACAGTGATCCATACAATAAAGA
 AGATGTTGAAGCTGTTGCGGTGGAGAAAGAAGAGAGTCGAAAAAGAGCAATAGGACAGTG
 TTCGGGCGTGGTGGCTCAAGACGCCAGTGATGTTTTAGGAGTTACAGAGTTAGAAGGAGA
 GGGTAAGAATGTTCTGTGAAGAGCCGAGAGTTTCAAGCTGATATGGAAA

>G622 Amino Acid Sequence (domain in AA coordinates: TBD)

MFEVKMGSKMCMNASCGTSTVWEKKGWPLRSGLLADLCYRCGSAYESSLFCEQFHKDQS
 GWRECYLCSKRLHCGCIASKVTIELMDYGGVGCSTCACCHQLNLNTRGENPGVFSRLPMK
 TLADRQHVNGESGGRNEGDLFSQPLVMGGDKREBFMPHRGFGKLMSPESTTTGHRLDAAG
 EMHESPLQPSLNMGLAVNPFSPSFATEAVEGMKHHISPSQSNMVHCSASNILQKPSRPAI
 STPPVASKSAQARIGRPPVEGRGRHLLPRYWPKYTDKEVQQISGNLNLNIVPLFEKTL
 ASDAGRIGRLVLPKACAEAYFPPISQSEGIPLKIQDVRGREWTFQFRYWPNNNSRMVLE
 GVTPCIQSMMLQAGDVTFTSRVDPGGKLIMSGSRKAANAGDMQCGCLTNGTSTEDTSSSGV
 TENPPSINGSSCISLIPKELNMGMPENLNSETNNGRIGDDPTRVKEKKRTRTIGAKNKRLL
 LHSEESMELRLTWEEAQDLLRSPSVKPTIVVIEEQEIEEYDEPPVFGKRTIVTTKPSGE
 QERWATCDDCSKWRRLPVDALLSFKWTCIDNVWDVSRSCSAPEESLKELENVLKVGREH
 KKRRTGERQAAQSQQEPCGLDALASAAVLGDTIGEPEVATTTTRHPRHRAGCSCIVCIQPP
 SGKGRHKPTCGCTVCSVKRRFKTLMRRKKKQLERDVTAAEDKKKMDLAESDKSKEE
 KEVNTARIDLNSDPYNKEDVEAFAVEKEESRKRAIGQCSGVVAQDASDVLGVTELEGEK
 NVREEPRVSS*

>G778 (50..1249)

TCTCAATAACACAAAACCTTTTAAACTAGTAAATACACAGATTTTAGGATGAGCCAATG
 TGTTCCAAACTGTCAATCGATGATACTCCGGCAGCAGCCACCACCCTCCGCTCCAC
 CACAGCCGAGACATCCCCATATTAGACTACGAGGTAGCCGAGCTGACGTGGGAGAACGG
 GCAACTAGGCTTGACCGCTTAGGTCCACCGCAGTGACGGCTTCGTGACCAAGTACTC
 CACAGGCGCCGGTGGAACTTGGAGTCGATAGTGGAACCAAGCTACTCGCTCCCTAAGCC
 TAAGCCACGATGAGCTCGTCCCGTGGTTCCATCATCGCTCCTCCAGGGCCGCGATGGC
 AATGGACGCGCTTGTCCTTGTCTCAACCTAGTACACGAGCAGCAGAGCAAGCCTGGTGG
 CGTTGGCTCCACCCGGGTGGGGTCATGTAGCGATGGTTCGTACCATGGGCGGTGGAAAACG
 AGCAAGAGTGGCACCCGAGTGGAGCGGCGGCGGAGTCAGCGGCTGACCATGGACACTTA
 CGACGTAGGTTTCACTCAACATCAATGGGCTCGCACGATAACACAATCGACGATCATGA
 CTCCGTCTGCCACACCGCCACAGATGGAGGACGAAGAAGAGAAGAAAGCCGGAGGAAA
 ATCATCAGTTTCAACCAAGAGAAGCAGAGCTGCTGCTATTTCATAACCAATCCGAACGTAA
 GAGGAGAGATAAAATCAATCAAAGGATGAAGACTTTGCAAAAACCTGGTTCCCAATTCAG
 CAAGACGGATAAAGCATCTATGTTGGATGAAGTGATAGAGTATTTGAAGCAACTTCAAGC
 ACAAGTGAGCATGATGAGCAGAATGAATATGCCTTCTATGATGCTTCTATGGCCATGCA
 GCAACAACAACAACTACAAATGTCTCTCATGTCCAATCCCATGGGTTTAGGGATGGGCAT

GGGGATGCCCCGCTCTCGGTCTCCTCGACCTTAATTCATGAACCGAGCTGCTGCAAGCGC
TCCTAATATCCATGCCAACATGATGCCAAACCCATTTTGGCCATGAATTGTCCATCGTG
GGATGCTTCTTCCAATGACTCTCGATTTTCAGTCTCCTCTCATCCCCGATCCTATGTCTGC
CTTTCTTGCATGCTCTACTCAGCCAACGACGATGGAAGCGTATAGCAGGATGGCTACATT
ATATCAGCAAAATGCAACAACAACCTTCTCCTCCTTCGAATCCAAAATGATTATTACTCAA
ACACCTCTATATAGTTTACGTCTATATATGTGTTAGTCACATACATATATATATATTC
CATCATAATTATTTATTTATATGTATAGGCTTCTCATGAATTATGATATTATACGTATTA
CGTAAAAAA

>G778 Amino Acid Sequence (domain in AA coordinates: 220-267)
MSQCVPNCHIDDTAAATTTVRSTTAADIPILDYEVAEELTWENGQLGLHGLGPPRVTASS
TKYSTGAGGTLESIVDQATRLPNPKPTDELVPWFHRRSSRAAMAMDALVPCSNLVHEQQS
KPGGVGSTRVGS CSDGR TMGGGKRARVAPEWSSGGGSQLTMDTYDVGFSTSTSMGSHDNTI
DDHDSVCHSRPQMEDEEEKKAGGKSSVSTKRSRAAAIHNQSERKRRDKINQRMKTLQKLV
PNSSKTDKASMLDEVIEYLKQLQAQVSMMSRMNMPMSMLPMAMQQQQQLQMSLMSNPMGL
GMGMGMPLGLLLDLNSMNRAAASAPNIHANMMPNPFPLMNCPSWDASSNDSRFQSPILPD
PMSAFLACSTQPTTMEAYS R MATLYQQMQQLPPPSNPK*

>G791 (173..877)
TTTTCTTTGGGTGTTCTTCCACCAACGGCAGAAATCGATTGGGCTTAAATCTCCCCCTC
CTTTCGATCTCTCTGATCGCCGCCGGGAACATTCAATTTCCCGGGAGTTCAACAAAAAA
AAACTCTCCGTTTTTATTTTTCCCTTTTTCACCGGTGGAAGTTTCCGGAGATGGTGTC
ACCCGAAAACGCTAATTGGATTTGTGACTTGATCGATGCTGATTACGGAAGTTTCAAT
CCAAGTCTCTGTTCTCTTGGCCTGTTTCAGCAACCTATTGGTGTTTCTTCTAACTCCAG
TGCTGGAGTTGATGGCTCGGCTGGAACTCAGAAGCTAGCAAGAAGCTGGATCCAAAAA
GAGGGGGAGATGTGAATCATCTCTGCCACTAGCTCGAAAGCATGTAGAGAGAAGCAGCG
ACGGGACAGGTTGAATGACAAGTTTATGGAATTGGGTGCAATTTTGAGCCTGGAAATCC
TCCCAAAACAGACAAGGCTGCTATCTTGGTTGATGCTGTCCGCATGGTGACACAGCTACG
GGGCGAGGCCCAGAAGCTGAAGGACTCCAATTCAAGTCTTCAGGACAAAATCAAGAGTT
AAAGACTGAGAAAAACGAGCTGCGAGATGAGAAACAGAGGCTGAAGACAGAGAAAGAAAA
GCTGGAGCAGCAGCTGAAAGCCATGAATGCTCCTCAACCAAGTTTTTCCAGCCCCACC
TATGATGCCTACTGCTTTTGGCTTCAGCGCAAGGCCAAGCTCCTGGAAACAAGATGGTGCC
AATCATCAGTTACCCAGGATTTGCCATGTGGCAGTTTCATGCCTCCTGCTTCAGTCGATAC
TTCTCAGGATCATGTCCTTCGTCCTCTGTTGCTTAATCAAGAAAAATCATCAACCGGTT
TGCTTCTTGCTTCCGCTTAAAGAAAAAGTCTCCATTGTGTTTGCTCTCCTCTCTTCTCG
GCTTTCTTAGTCTTATCCTTTTGTCTTGTGCTGTTATCATCGTAACTGTTATCTGTTGAA
CAATGATATGACATTGTAACTCCAATTGCTTCGCGCAATGTTATCTATTACATGTAAA
TTTAAGTAGAGTTTGGCAAAAAAAA

>G791 Amino Acid Sequence (domain in AA coordinates: 75-143)
MVSPENANWICDLIDADYGSFTIQPGFSWVPVQQPIGVSSNSSAGVDGSAGNSEASKEPG
SKKRGRCESSSATSSKADREKQRRDLNDKFELGAILEPGNPPKTDKAAAILVDVRMVT
QLRGEAQKLKDSNSSLQDKIKELKTEKNELRDEKQRLKTEKEKLEQQLKAMNAPQPSFFP
APPMPPTAFASAQGPAGNKMVPIISYPGVAMWQFMPPASVDTSQDHVLRPPVA*

>G861 (158..880)
CTTCTCTCCTCCTCCATCTCTTCTCTTTACTCTCTCTTTAATCATCTCTCATTCTTGA
ATCTTGATCCATCAAAATCAATCCCGTTCTCGAAAGATCCATTAATCAAAACCTAAGC
TCTCTCTCTTGCTTCTAGGGTTTTTTTGTTCGTTGTGATGGCGAGAGAAAAGATTGAGAT
CAGGAAGATCGACAACGCAACGGCGAGACAAGTGACGTTTTTCGAAACGAAGAAGAGGGCT
TTTCAAGAAAGCTGAAGAACTCTCCGTTCTCTGCGACGCCGATGTGCTCTCATCATCTT
CTCTTCCACCGGAAAAGTTTCGAGTTCTGTAGCTCCAGCATGAAGGAAGTCTTAGAGAG
GCATAACTTGCGAGTCAAGAACTTGGAGAAGCTTGATCAGCCATCTCTTGAGTTACAGCT
GGTTGAGAACAGTGATCACGCCCCGAATGAGTAAAGAAATTGCGGACAAGAGCCACCGACT
AAGGCAAATGAGAGGAGAGGAACCTCAAGGACTTGACATTGAAGAGCTTCAGCAGCTAGA
GAAGGCCCTTGAAGCTGGTTTGACGCGTGTGATTGAAACAAAGAGTGACAAGATTATGAG
TGAGATCAGCGAAGTTTCAGAAAAAGGGAATGCAATTGATGGATGAGAACAAAGCGTTGAG
GCAGCAAGGAACGCAACTAACGGAAGAGAACGAGCGACTTGGCATGCAATATGTAACAA
TGTGCATGCACACGGTGGTGTGTAATCGGAGAACGCTGCTGTGTACGAGGAAGGACAGTC
GTCGGAGTCTATTACTAACGCCGGAAGCTTACCGGAGCGCTGTTGACTCCGAGAGCTC
CGACACTTCCCTTAGGCTCGGCTTACCGTATGGTGTTAGAGATGGAACAATCAAGAA

GTGATGGAGTGAGGAGAGTAATGTAAATCTTTTAACTCGGTAGTAACAAGAGACAATG
TCTAAGTAGTGAATTCTCAAATGTTTGTGTAAGTTTCTGCCTATGGAAGAGGCTTTCATT
TTTATGATTTTCACTATGTATGATCTCTTCACTGCATTTCTGGTTAGTAACGGCTTGT
CACCGATAAATTTCTCGTTATGGAAAGTTAGAATAAAAAAAAAAAAAAAAAAAAAA
>G861 Amino Acid Sequence (domain in AA coordinates: 2-57)
MAREKIQIRKIDNATARQVTFKRRRLFKKAEELSVLCDADVALIIFSSTGKLFECSS
SMKEVLERHNLQSKNLEKLDQPSLELQLVENS DHARMSKEIADKSHRLRQMRGEELQGLD
IEELQQLEKALETGLTRVIETKSDKIMSEISELQKKGMQLMDENKRLRQQGTQLTEENER
LGMQICNNVHAHGAESENAAYVEEGQSSESITNAGNSTGAPVDSSESDTSLRLGLPYGG
*

>G938 (1..1755)

ATGATGATGTTTAAACGAGATGGGAATGTATGGAAACATGGATTTCTTCTCTTCTCCACA
TCTCTCGATGTGTGCCATTACCACAAGCTGAACAAGAACCTGTAGTTGAAGATGTGCAC
TACACCGATGATGAGATGGATGTGGATGAGCTTGAGAAGAGGATGTGGAGAGACAAAATG
CGTTTGAACGCTCTCAAGGAGCAACAGAGTAAGTGTAAGAAGGCGTCGATGGTTCGAAA
CAGAGGCAGTCGCAAGAGCAAGCTAGGAGGAAGAAAATGTCTAGAGCCCAAGATGGGATC
TTGAAGTATATGTTGAAGATGATGGAAGTTTGTAAAGCTCAAGGCTTTGTTTATGGTATT
ATTCTGAGAAGGGTAAGCCTGTGACTGGTGCTTCGGATAATTTGAGGGAATGGTGGAAA
GATAAGGTTAGGTTTGTATCGTAATGGTCCAGCTGCTATTGCTAAGTATCAGTCAGAGAAT
AATATTCTGGAGGGAGTAATGATTGTAACAGCTTGGTTGGTCCAACACCGCATACGCTT
CAGGAGCTTCAGGACACGACTCTTGGTTTCGCTTTTATCGGCTTTGATGCAACATTGTGAT
CCACCGCAGAGACGGTTTCTTTGGAGAAAGGAGTTTCTCCACCTTGGTGGCCTAATGGG
AATGAAGAGTGGTGGCCTCAGCTTGGTTTACCAAATGAGCAAGGTCCTCCTCCTTATAAG
AAGCCTCATGATTGAAGAAAGCTTGGAAAGTCGGTGTTTAACTGCGGTGATCAAGCAT
ATGTCGCCGGATATTGCCGAAGATCCGTAAGCTTGTGAGGCAATCAAAATGCTTGCAGGAT
AAGATGACGGCGAAAGAGAGTGCTACTTGGCTTGCCATTATTAACCAAGAAGAGGTTGTG
GCTCGGGAGCTTTATCCCGAGTCATGCCCTCCTCTTTCTTCTTCTTCTCATTAGGAAGC
GGGTCGCTTCTCATTAAATGATTGTAGCGAGTATGACGTTGAAGGTTTCGAGAAGGAACAA
CATGGTTTCGATGTGGAAGAGCGGAAACCAGAGATAGTGATGATGCATCCTCTAGCAAGC
TTTGGGGTTGCTAAAATGCAACATTTTCCCATAAAGGAGGAGGTCGCCACCACGGTAAAC
TTAGAGTTCACGAGAAAGAGGAAGCAGAAACAATGATATGAATGTTATGGTAATGGACAGA
TCAGCAGGTTACACTTGTGAGAATGGTCAGTGTCTCAGCAAAAATGAATCTTGGATTT
CAAGACAGGAGTTCAAGGGACAACCAACAGATGGTTTGTCCATATAGAGACAATCGTTTA
CGTATGGAGCATCCAAGTTTCAATATGGGTGGAATGAAACTAGTAGTTCCTCAGCAACCA
GTCCAACCGATCGACCTATCGGGCGTTGGAGTTCGGGAAAACGGGCAGAAGATGATCACC
GAGCTTATGGCCATGTACGACAGAAATGTCCAAAGCAACCAACGCCTCCTACTTTGATG
GAAAACCAAAGCATGGTCATTGATGCAAAAGCAGCTCAGAATCAGCAGCTGAATTCAAC
AGTGGCAATCAAATGTTTATGCAACAAGGGACGAACAACGGGGTTAACAATCGGTTCCAG
ATGGTGTGTTGATTTCGACACCATTCGATATGGCAGCATTTCGATTACAGAGATGATTGGCAA
ACCGGAGCAATGGAAGGAATGGGGAAGCAGCAGCAGCAGCAGCAGCAGCAAGATGTA
TCAATATGGTTCTGA

>G938 Amino Acid Sequence (domain in AA coordinates: 96-104)
MMMFNEMGMYGNMDFSSSTSLDVCPLPQAEQEPVVEDVDYTDDEMDVDELEKRMWRDKM
RLKRLKEQQSKCKEAGVDSKQKQSQEQARRKMSRAQDGILKYMLKMMEVCKAQGFVYGI
IPEKGPVTGASDNLRWWKDKVRFDRNGPAAIAKYQSENNISGGSNDNCNSLVGPTPHTL
QELQDITLGSLLSALMQHCDPPQRRFPLEKGVSPWPWPNGNEEWPPQLGLPNEQGPYPYK
KPHDLKKAWKVGVLTAIVIKHMSPDIAKIRKLVRQSKCLQDKMTAKESATWLAIINQEEVV
ARELYPESCPPLSSSSSLGSGSLLINDCSEYDVEGFEKEQHGFDVEERKPEIVMMHPLAS
FGVAKMQHFPIKEEVATTNLEFTRKRKQNDNMVMVMDRSAGYTCENGQCPHSMNLGF
QDRSSRDNHQMVCFYRDNRLAYGASKFHMGMKLVVPQQPVQPIDLSGVGVPENGQKMIT
ELMAMYDRNVQSNQTPPTLMENQSMVIDAKAAQNQLNFNSGNQMFQQGTNNGVNRRFQ
MVFDSPTFDMAAFDYRDDWQTGAMEGMGKQQQQQQQQQDVSIFW*

>G965 (73..1956)

GATTCTCTGTGTATGTCTGAATCCTTACAGGATCCAAGAGCTTTGGAAAAAGATATAAT
GAATAACAAGATATGGGTTTAGCTACTACAACTTCTTCTATGTACAAAGATTATCATCAT
CACCAAGGAATCTTTCTTCTCTAATGGATTCCACCGATCATCATCAACCACTCATCAG
GAGGAAGTAGATGAATCCGCCGTCGTCTCCGGTGCTCAAATTCGGGTTTATGAAACCGCC

GGAATGTTGTCTGAAATGTTTGCTTACCCTGGCGGAGGTGGCGGCGGTTCCGGTGGAGAG
 ATTCTTGATCAGTCTACTAAACAGTTGCTAGAGCAACAAAACCGTCACAACAACAACAT
 AACTCAACTCTTCATATGTTATTACCAAATCATCATCAAGGTTTTGCTTTCACCGACGAA
 AACACTATGCAGCCGAGCAACAACAACACTTTACATGGCCATCTTCTCCTCCGATCAT
 CATCAAAACCGAGATATGATCGGAACCGTCCACGTGGAAGGAGGAAGGGTTTGTCTTTA
 TCTCTCTCATCTTCATTAGCCGAGCTAAAGCCGAGGAATATAGAAGCATTATTGTGCA
 GCCGTTGATGGAACCTTCTTCTTCTTAACGCATCCGCTCATCATCAATTCAATCAG
 TTCAAGAATCTTCTTCTTGAGAATTCTTCTTCTCAACATCATCACCATCAAGTTGTGGGA
 CATTTTGGTTCATCATCATCATCTCCCATGGCGGCTTCTTCATCCATTGGAGGGATCTAC
 ACGTTGAGGAATTGCAAAATATACGAAACCGGCTCAAGAGTTGTTGGAAGAGTTTTGTAGT
 GTTGAAGAGGACATTTCAAGAAGAACAACCTTAGTAGGAACAACCTCAAACCCTAATACT
 ACCGTTGGAGGAGGAGGCGGAGGGTCTCGTCAATCGGCCGGAACAGCTAATGATAGTCCT
 CCTTTGTCTCCGGCTGATCGGATTGAACATCAAAGAAGAAAAGTCAAGCTACTATCTATG
 CTTGAAGAGGTGGACCGACGGTACAACCACTACTGCGAACAATGCAAATGGTAGTGAAC
 TCATTGACCAAGTAATGGGTACGGCGCGCGGTTCCGTACACGACATTAGCTCAAAAG
 GCAATGTCTAGGCATTTCCGGTGTGTTGAAAGACGCGGTAGCGGTTTCAAGTTAAACGCGAG
 TGTGAGCTTCTAGGGGATAAAGAGGCGGCGAGGGCTGCATCCTCGGGGTTAAACAAAGGG
 GAAACGCGCGGATTGCGTTTGTCTAGAGCAGAGTTTGCCTGAGCAACGAGCGTTTCATCAT
 ATGGGTATGATGGAGCAAGAGGCATGGAGACCGCAACGTGGTTTGCCTGAACGCTCCGTT
 AATATCCTTAGAGCTTGGCTATTTCGAGCATTTTCTTAATCCGTACCCAAGCGATGCTGAT
 AAGCACTCTTAGCAGCAGACTGGTTTATCCAGAAATCAGGTGTCAAATTGGTTCATA
 AATGCTAGGGTTTCGCTATGGAACCAATGGTGAAGAGATGTATCAACAAGAAGCAAAA
 GAAAGAGAAGAAGCAGAAGAAGAAAATGAAAATCAACAACAACAAGAAGACAGCAACAA
 ACAACAACAACGACAGCAAAACCAACAACAATGAAAACAACCTTACTGTCTATAACCGCA
 CAAACTCCAACGACGATGACATCGACACATCACGAAAACGACTCTTCATTCTCTCTTCC
 GTCGCCCGCGCTTCTCACGGCGGTTCAGACGCGTTCAACGTCGCCACGTGTCAGCAAGAC
 GTCAGTGACTTCCACGTCGACGGAGATGGTGTGAACGTCATAAGATTCGGGACCAACAG
 ACTGGTGACGTGTCTCTTACGCTTGGTCTACGCCACTCTGGCAATATCTCTGATAAGAAC
 ACTTCTTCTCEGTTAGAGACTTTGGAGATTTTTAGTCTTCTTTGTTTCTCAATTTATTC
 ATC

>G965 Amino Acid Sequence (domain in AA coordinates: 423-486)

MGLATTTSSMSQDYHHHQGIFSFNNGFHRSSSTTHQEEVDESADVSGAQIPVYETAGMLS
 EMFAYPQGGGGGGSGGEILDQSTKQLLEQQNRHNNNNNSTLHMLLPNHHQGFATDENTMQ
 PQQQQHFTHWPSSSSDHHQNRDMIGTVHVEGGKGLSLSLSSSLAAAKAEFYRSIYCAAVDG
 TSSSSNASAHHQFNQFKNLLLENNSSQHSHHHQVVGHFSSSSSPMASSSIGGIYTLRN
 SKYTKPAQELLEEFCSVGRGHFKNKLNRNNSNPNTTGGGGGGSSSSAGTANDSPPLSP
 ADRIEHQRRKVKLLSMLEEVDRRYNHYCEQMOMVNSFDQVMGYGAAPVYTTLAQKAMSR
 HFRCLKDAVAVQLKRSCELLGDKEAAGAASSGLTKGETPRLRLLEQSLRQRAFHHMGM
 EQEAWRPQRLPERSVNILRAWLFEHFLNPYPSDADKHLARQTGLSRNQVSNWFINARV
 RLWKPMVEEMYQOEAKEREAEENENQQQORRQQQTNNNDTKPNNNNNNFTVITAQTPT
 TMTSTHHENDSSFLSSVAAASHGGSDAFTVATCQQQDVSDFHVDGDGVNVIRFGTKQTGDV
 SLTLGLRHSNIPDKNTSFSVRDFGDF*

>G1143 (54..677)

AAATAAGAAATATAAACACTTTTGTCTGAAAAATTATCAAAGAAGAAGAAATAAATGGGTG
 GAGGAAGCAGATTTCAAGAACCAGTGAGGATGAGCCGTAGGAAACAAGTAACAAAAGAGA
 AGGAAGAAGATGAAAACCTTCAATCTCCAAATCTTGAAGCAGAGAGACGTAGAAGAGAGA
 AGCTTCATTGTGCGCTTATGGCTCTGCGATCTCATGTCCCCATTGTACCAACATGACTA
 AAGCAAGTATTGTTGAAGATGCGATTACTTACATAGGAGAGCTTCAAAACAATGTTAAGA
 ATCTCTTAGAGACATTTATGAAATGGAAGAAGCTCCTCCTGAGATTGATGAAGAACAAA
 CGGATCCAATGATAAAACCTGAAGTTGAAACTAGTGATCTTAACGAAGAGATGAAGAAAC
 TCGGAATCGAGGAGAATGTGCAATTGTGTAAGATTGGGGAGAGGAAGTTTTGTTTAAAGA
 TCATAACAGAGAAGAGAGATGGGATCTTTACTAAATTCATGGAGGTTATGAGATTTCTCG
 GATTGAGATTATCGATATTAGTCTAACAACCTTCAAATGGAGCAATTCTTATTAGTGCCT
 CTGTTTCAAGACAGGAACCTCTGTGATGTTGAACAGACAAAAGATTTTCTTTTGAAGTTA
 TGAGAAGCAATCCATAAGTATTAAATATATACATCTTGGAATTTCTTGATCTAATAACA
 TTTCCATTGGTTTTTATTACATTGTTGTTCCATTTTAAATATGATATGATTGAGATGAAA
 AAGAGTTTGTGTTACAAGCCAATGA

>G1143 Amino Acid Sequence (domain in AA coordinates:33-82)
MGGGSRFQEPVRRSRKQVTKEKEEDENFKSPNLEAERRRRREKLHCRLMALRSHVPIVTFN
MTKASIVEDAITYIGELQNNVKNLLETFHMEEEAPPEIDEEQTDPMIKPEVETSDLNEEM
KKLGIEENVQLCKIGERKFWLKIITEKRDGIFTKFMEVMRFLGFEEIIDISLTTSNGAILI
SASVQTQELCDVEQTKDFLLEVMRSNP*

>G1190 (209..2020)
TCCTGTCCCAAAACCAAGACTTGAGAGTGTGTCTTTAGAGAGAGATCTTCTCTCTTTT
ATCTTACGACTCTCACTTCTTATCTCAAATCTACTTCAACTCTATTTCCAGTCTCCACAT
TTTCCACAAATTTCAACTCTTGTCTCTTCTCCAAAGTAAAAACAAATCGTTGCAAG
TGAGGTTTGGTTTTGGTGTATAGAATTATGAAGAGCGGGAAGCAATCTTCGCAACCTGA
AAAGGGTACTTCCAGGATCTTGTCACTGACTGTCTGTATTATCGCATTTTGCCTTTCTC
CTTCTACCTCGGTGGTATATTTTGTCTCTGAGAGAGACAAGATTGTAGCCAAGGATGTCAC
AAGGACGACTACAAAGGCTGTAGCTTCCCTTAAAGAACCTACAGCTACTCCTATTCAAAT
CAAATCCGTTTCTTTCCCGGAGTGCCTGGTCCAGAGTTCCTCAAGATTACACCCCGTGCACCGA
TCCAAAGAGGTGGAAGAAGTATGGTGTCCATCGCTTAAGTTTCTTGGAGCGTCATTGTCTC
TCCGGTATATGAAAAGAATGAGTGTGATTCCACCACCAGACGGGTATAAACCGCTAT
AAGATGGCCCAAGAGCCGAGAACAGTGTGTTGGTACAGGAACGTGCCTTATGATTGGATCAA
TAAGCAAAAGTCTAACCAGCATTTGGCTTAAGAAAGAAGGAGATAAGTTCCATTTCCCTGG
TGGTGGTACCATGTTCCCTCGTGGAGTTAGTCACTATGTTGATTGATGCAAGATCTGAT
TCCTGAAATGAAAGACGGAACAGTCAGGACCGCCATTGATACTGGCTGTGGGGTTGCGAG
CTGGGGAGGCGATCTTTTGGACCGTGGGATACTATCACTCTCTCTTGGCTCCAAGAGATAA
CCATGAAGCTCAGGTTCAATTTGCTCTTGAACGTGGAATTCCTGCGATTCTCGGGATCAT
CTCTACGCAACGTCTCCCTTTTCTTCAAATGCATTTGATATGGCTCATTGTTCAAGATG
TCTTATTTCCCTGGACAGAATTTGGTGAATCTATTTACTTGAGATTCACCGTATAGTTCCG
ACCTGGAGGTTTTTGGGTTCTTTTCTGGTCCACCTGTGAACTATAATAGACGATGGCGTGG
ATGGAACACCAACCATGGAAGATCAGAAATCTGACTACAACAAGCTTCAGTCACTTCTAAC
CTCCATGTGTTTCAAAAAGTACGCTCAAAAAGATGACATAGCCGTGTGGCAGAACTCTC
AGACAAATCTTGCTATGACAAAATCGCTAAGAATGGAAGCTTACCCTCCCAATGTGA
CGACAGTATAGAACTGATTCTGCTTGGTACACTCCACTCCGTCTTGCCTGGTGGTGGCCCC
GACACCTAAAGTCAAGAAGTCTGGTCTCGGATCAATCCCAAAATGGCCCCGAGAGGTTACA
TGTCGCGCCCGAGAGAATCGGTGATGTTACCGGAGGGAGTGCAGAACAGTTTGAAACACGA
TGATGGTAAATGGAAGAACAGAGTTAAGCATTACAAGAAAGTTTACCAGCTCTTGGGAC
AGACAAGATAAGAAATGTTATGGATATGAACACTGTTTATGGAGGTTTCTCTGCGGCCCT
CATTTGAGGATCCCATTTTGGGTCATGAACGTTGTATCATCGTACAGCGCAAATTCGCTTCC
TGTTGTCTTTGATCGCGGTCTCATCGGGACTTACCACGACTGGTGCAGAGCTTTCTCAAC
GTATCCAAGACATATGATCTTCTTACCTCGACAGTCTTTTTACCTTGGAGAGTCAACAG
GTGTGAGATGAAGTACATTTTGTAGAGATGGACAGGATCTTGCAGCCGAGTGGATATGT
TATAATCCGAGAATCGAGTTATTTTATGAGCGCAATCACAACGTTAGCGAAAGGGATAAG
GTGGAGTTGCCGAGAGAGGAGACTGAGTATGCAGTCAAAAGTGAGAAGATTCTGGTTTG
CCAGAAAAAGCTATGGTTTTCGTCAAACCAAACCTCTTGATGAGACCACCTGTATCATAG
TGTTTATCATCTCCTGTGATGCACACTACAGAGAGAAGGATCTAGTCCTTTGAGTCCAAG
ATATAGCTCTATAACAAATCTCCTTTTTTTTGTCTCTTTAATTTCTTGGGTATTTTACCGG
TATAGATTGATATTATATATTTTTTAATTATATTTTAAATATATAGATATATTAGTATGT
GGTTTAAACACTATTATTATCAAGGTCTTAAAGATTTGCTTTGCAAGAGTTAAAAATGT
TGGAGTAAGGACCTCTTGATTAATAAATTGACTGACGCAGCAA

>G1190 Amino Acid Sequence (domain in AA coordinates: entire protein)
MKSGKQSSQPEKGTSLRLSLTVLFIAFCGFSFYLGIGFCSEKIVAKDVTRTTTKAVAS
PKEPTATPIQIKSVSFPECGSEFQDYTPCTDPKRWKYGVHRLSFLERHCPPVYEKNECL
IPPPDGYKPIRWPKSREQCWYRNPYDWINKQSNQHWLKKEGDKFHPGGGTMPFRGV
SHYVDLMQDLIPEMKDGTVRTAIDTGCGVASWGGDLLDRGILSLAPRDNHEAQVQFAL
ERGIPAILGIISTQRLFPSPNAFDMAHCSRCLIPWTEFGGIYLLIHRIVRPGGFVWLSG
PPVNYNRRWRGWNTTMEDQKSDYNKLQSLTSMCFKKYAQKDDIAVWQKLSKSCYDKIA
KNMEAYPPKDDSI EPDSAWYTPLRPCVVAPTPKVKKSGLGSIPKWPERLHVAPERIGDV
HGGSANSLKHDDGKWKNNRVKHYKVLPA LGTDKIRNVMDMNTVYGGFSAALIEDPIWVMN
VVSSYSANSLPVVDFDRGLIGTYHDWCEAFSTYPRTYDLLHLSLFTLESHRCMKYILLE
MDRILRPSGYVIIRESSYFMDAITTLAKGIRWSCRRETEYAVKSEKILVCQKKLWFSSN
QTS*

>G1198 (230..1675)

TCTTTTCAAATTCCAATCATTTGATCAACTAATCAAGAATTAATTATAAGACTTTGCAAT
CTCTCTCCCTCTCCCTCTCCCTAGCTAGTTCTCTCTTGTGTTTCTTAACCTCGAGCTTCTC
TCAATAGTGATTATCATCTTTTTTCATCATTTCAAGATTTAATGTGTTTTGCAGAAAAGAG
ACTAATCAAGAAGAGATATCATCAATTGAAGCTGTTTTCTTGAGTAGAGATGGCGAACCA
TAGAATGAGCGAAGCTACAAACCATAACCACAATCATCATCTTCCTTATTCACTTATTCA
TGGTCTCAACAACAATCATCCATCTTCTGGTTTCATTAACCAAGATGGATCGTCCAGTTT
CGATTTTGGAGAGCTAGAAGAAGCAATTGTTCTGCAAGGTGTCAAGTATAGGAACGAGGA
AGCCAAGCCACCTTTATTAGGAGGAGGAGGAGGAGCTACGACTCTGGAGATGTTCCCTTC
GTGGCCAATCAGAACTACCAAACCTTCTCTACTGAGAGTTCCAAGTCAGGAGGAGAGAG
CAGCGATTCAAGGATCGGCTAATTTCTCCGGCAAAGCTGAAAGTCAACAACCGGAGTCTCC
TATGAGTAGCAAAACATCATGCTTCAACCTCATCATAATAACATGGCAAACCTCAAG
TTCAACATCTGGACTTCTTCCACTTCTCGAAGTTTAGCTCCTCTAAACCTTCGGAAGA
TAAGAGGAAGGCTACAACCTCAGGCAAAACAGCTTGATGCTAAGACGTTGAGACGTTTGGC
CCAAAATAGAGAAGCTGCTCGCAAAAGCCGTCTTAGGAAAAAGCGTATGTGCAACAGCT
AGAATCAAGTAGGATAAAGCTTTCCCAATTGGAGCAAGAACTTCAGCGAGCTCGTTCTCA
GGGGCTGTTTATGGGTGGTTGTGGACCACCAGGACCTAACATCACTTCCGGAGCTGCAAT
ATTTGACATGGAATATGGGAGATGGCTAGAGGATGATAACCGGCATATGTGCGGAGATTG
AACCGGTCTTCAGGCTCATTTATCTGACAATGATTAAAGGTTGATCGTTGACGGTTACAT
TGCTCATTTTGTATGAGATATCCGATTAAAAGCCGTGGCAGCGAAAGCCGATGTTTTTCA
CCTCATCATTGGGACATGGATGTCCCCAGCCGAACGTTGTTTTATTGGATGGCTGGTTT
CCGTCCATCCGACCTAATCAAGATATTGGTGTGCGAAATGGATCTATTGACGGAGCAACA
ACTGATGGGAATATATAGCCTACAACACTCGTTCGCAACAAGCAGAGGAGGCTCTCTCGCA
AGGCCTCGAACAACCTCAGCAATCTCTCATCGATACTCTCGCCGCACTCTCCAGTCATTGA
CGGAATGCAACAAATGGCTGTGCTCTCGGAAAGATCTCTAATCTCGAAGGCTTTATCCG
CCAGGCTGATAACTTGAGGCAGCAGACCGTTACCAGCTGAGGCGGATCTTGACCGTCCG
ACAAGCTGCACGGTGTTCCTAGTCATCGGAGAGTACTATGGACGGCTCAGAGCTCTTAG
CTCCCTTTGGTTGTACGCCCCAGAGAGACTGATGAGTGATGAAACCTCTTGTCAAAC
GACGACGGATTTCAGATTGTTTCAGTCATCTCGGAACCACTTCTCCAATTTCTGAATGGA
ATGAAACTTTGTATAAATAAAGGCCAAGTTTCATTGTCTGTGCGTAATTTACCTATTTT
CTTTAAAGTTGTACTAGAGAAAAGATAGGATCTTCTCTTCG

>G1198 Amino Acid Sequence (domain in AA coordinates: 173-223)
MANHRMSEATNHNHNNHLPYSLIHGLNHNHPSSGFINQDSSSFDGEELEEAIVLQGVKY
RNEEAKPPLLGGGGGATTLEMFPSWPIRTHQTLPTSSKSGGESSDSGSANFSGKAESQQ
PESPMSSKHLMQLPHNNMANSSSTSGLPSTSRLLAPPKPSKEDKRKATTSKGQLDAKTL
RRLAQNRREARKSRLRKKAYVQQLSSRIKLSQLEQELQARSQGLFMGGCGPPGPNITS
GAAIFDMYGRWLEDDNRHMSEIRTLQAHLSNDLRLIVDGYIAHFEIIFRLKAVAAGA
DVFHLIIGTWMSPAERCFIWMAGFRPSDLIKILVSQMDLLTEQQLMGIYSLQHSQQABE
ALSQGLEQLQQLIDTLAASPVIDGMQMAVALGKISNLEGFIRQADNLRQQTVHQLRRI
LTVRQAARCFIVIGEYYGRLRALSSLWLSRPRETLMSDETSCQTTTDLQIVQSSRNHFSN
F*

>G1226 (212..1159)

CTGCATTTATTAAGAACAGTTTAGAAAAGTGTCAACCCCTAAAGGAATGTTTTTAGTTTAG
AGGAAAGAGAGAGAAGAAGAGCAGCAGCAGAAGTTGTTAATTTGAAGACTATTTGAGGA
AAGACACCTATATCTAAATACTCAAAGTTACAAAAATATTACTTCAGAAAAAGTTCCAT
TAGAGAGACTCATAAAGCTTCTCATCTAATTATGAGTGGATTGATGAGTTTGGTGAATT
AGAAGACCAATTTGGTCAGATTTTCAGACACTACTATGGAAGAGAAGATACCATTTCTGCA
AATGCTTCAATGCATAGAACACCCCTTTTACAACAACAGAACCAATCAGTTTCTCCAATC
ACTTCTCCAGATCCAAACCTTAGAATCAAAGAGCTGTCTCACCCCTTGAAACAAACATCAA
AAGAGATCCGGGTCAAACAGATGACCCGGAAGGATCCAAGAACAGAAAACGGAGCAGT
AACGGTCAAAGAAAAAGAAAACGGAAACGTACAAGAGCTCAAAGAAACAAAGACGAAGT
TGAAAACCAAAGGATGACTCACATTGCGGTGCAACGTAATCGAAGACGACAAATGAACGA
ACACTTAACTCTCTCGATCTCTCATGCTCCTTCGTTTCTTCAACGGGGTGACCAAGC
TTCGATTGTAGGAGGGGCAATAGATTTCATCAAGGAAGTAGAGCAACTCTTGCAATCTCT
AGAAGCTGAGAAACGAAAGGATGGAAGTGAAGTCTTAAACGGCGTCTGTTCTTCT
ATCTTCGTCTCTTGCATGCACTAACTCTTCTATTCTTAGCGTGTCTACGACGTGCGAAAA
TGGATTTACGCGGAGATTGCGCGGTGGAGATACGACAGAAGTGGAGGCTACGGTGATACA

GAACCATGTGAGCTTAAAAGTTCCGGTGTAAAGAGAGGAAAACGACAGATCTTAAAAGCTAT
TGTCTCGATTGAAGAACTAAAGCTTGCATTCTACATCTCACTATCTCTTCTTCCCTTGA
CTTTGTCTACTACTCTTTCAATCTCAAGATGGAAGATGGTTGTAAATTAGGATCAGCAGA
TGAGATAGCGACAGCCGTTTCATCAGATCTTCGAGCAAATCAACGGTGAAGTCATGTGGTC
AAATCTTAGTCGAACCTAGTTGACTTTTGACTCCTAGTAACGTGTGTAAACTTTAGGTTA
CAAAGAAAAGGGACGTGATATAAATAAGAAAAACCAAAGAGGTGAAATTTTGGGAGTTTT
AATTATTATCTTATACTTTTTGGATTTTAGATTAGTAGCAAACCTCGCAGTGTCTACGAT
GACATTATTATTGGTCCATGAAGGTTTAGGTTAAAAA
>G1226 Amino Acid Sequence (domain in AA coordinates:115-174)
MSGLMSFGELEDQFGQISDTTMEEKIPFLQMLQCIEHPFTTTEPNQFLQSLLQIQTLESK
SCLTLETNIKRDPGQTDDEKDPRTENGAVTVKEKRKRKRTRAPKNKDEVENQRMTHIAV
ERNRRRQMNHLNLSRLMPPSFLQRGDQASIVGGAIDFIKELEQLLSLEAEKRKDGTD
ETPKTASCSSSSSLACTNSSISSVSTTSSENGFTARFGGGDTTEVEATVIQNHVSLKVRCK
RGKRQILKAIVSIEELKLAILHLTISSSFDFVIYSFNLKMEDGCKLGSAD E IATAVHQIF
EQINGEVMWSNLSRT*
>G1451 (124..2559)
TTTGTACTTCCGGAGCTAAAGAGTTATAGCTACTGTAGTAGCTGGAAGTGAAGAAGATTT
TTTAATAGATTGTACGGAATAATAGGGTTTTCAAAGTTTGGTTTCTTGAAGTTGAATTA
GACATGAAGCTGTCAACATCTGGATTGGGTCAACAGGGTCATGAAGGAGAGAAGTGTCTG
AATTCTGAGCTATGGCATGCTTGTGCTGGACCATTAGTCTCTCTTCCATCATCTGGTAGT
CGAGTTGTTTTACTTTCCACAGGGTCACAGTGAACAGGTAGCTGCTACAATAAAGGAA
GTTGATGGTCACATACCCAATTACCCAAGCCTACCACCACAATTGATATGCCAGCTCCAT
AATGTTACAATGCATGCAGATGTTGAGACGGATGAAGTCTATGCTCAAATGACACTTCAA
CCATTGACACCGGAGGAGCAGAAGGAAACATTTGTACCGATTGAGTTGGGGATACCGAGT
AAGCAACCTAGTAATTATTTTTTGAAGACTCTCACAGCTAGTGATACCAGTACACATGGA
GGGTTTTCTGTTCTAGACGTGCTGCTGAGAAAGTGTTCCTCCATTGGATTACACACTG
CAGCCACCAGCTCAAGAACTGATTGCAAGGGATCTCCATGATGTTGAATGGAAGTTTAGG
CATATCTTTTCGGGGACAGCCCAAACGGCATCTCCTAACTACTGGATGGAGTGTCTTTGTC
AGTGCCAAGCGACTAGTAGCTGGAGATTCTGTCAATTTTCATCAGGAATGAAAAGAATCAA
CTCTTTTTTGGGAATTTCGTATGCCACTCGGCCGAGACTATTGTACCATCATCTGTTTTA
TCTAGTGATAGCATGCATATTGGACTCCTTGCTGCTGCTGCACATGCTTCTGCAACTAAT
AGCTGTTTCACTGTTTTCTTTCATCCAAGGGCTAGCCAATCTGAGTTTGTGATACAATTT
TCCAAGTACATTAAAGCCGTTTTTTCACACGCGTATTTTCAGTTGGGATGCGTTTTCGCATG
CTCTTCGAGACAGAAGAGTCGAGTGTCCGCGAGTACATGGGTACTATAACTGGTATTAGT
GATCTAGATTCTGTTTCGTTGGCCAACTCTCATTGGCGATCTGTGAAGGTTGGTTGGGAT
GAATCGACTGCAGGGGAGAGACAGCCAAGGGTTCTTTATGGGAGATTGAGCCTCTGACT
ACCTTTCCTATGTATCCATCTCTTTTTCTCTCAGACTAAAACGTCCATGGCATGCTGGC
ACATCATCTTTGCCTGATGGAAGGGGTGATTTGGGAAGTGGTCTAACATGGCTAAGAGGG
GGAGGTGGAGAGCAGCAAGGTTTGCTTCTCTAAATTATCCATCTGTTGGTTTGTTCCTCA
TGGATGCAACAAAGGCTGGATCTCAGTCAAATGGGGACTGATAATAATCAGCAATACCAA
GCAATGTTAGCTGCTGGGTTGCAGAACATCGGCGGTGGAGATCCTTTAAGACAGCAGTTT
GTACAGCTGCAAGAGCCTCACCACCAATATCTTCAACAATCAGCTTCCCATTAATTCTGAT
TTGATGCTTCAGCAGCAACAGCAGCAACAAGCGTCACGCCATCTCATGCATGCTCAAACA
CAGATTATGAGTGAGAATCTCCGAGCAGAATATGCGACAAGAAGTTAGTAACCAACCA
GCTGGACAGCAGCAACAGCTACAGCAACCGGACCAAAATGCATATCTTAATGCTTTCAAA
ATGCAAAATGGCCATCTTCAACAGTGGCAGCAGCAATCAGAGATGCCATCTCCCTCGTTC
ATGAAGTCAGATTTTACTGACTCAAGCAACAAATTTGCAACAACCTGCTAGTCCGGCTTCT
GGAGATGGCAATCTTTTGAATTTTCTATAACCGGTCAGTCTGTACTCCCTGAGCAGTTA
ACAACAGAGGGCTGGTCTCCAAAGCATCCAACACTTTTCTGAACCGTTGTCACTTCCA
CAAGCCTATCTGGGAAGAGTCTTGCTCTAGAACCCGGAATCCGAGAAATCCCTCTCTT
TTCGGTGTGATCCCGACTCTGGACTCTTCTCCTCCCGAGTACGGTCCCGCTTTGCTTCT
TCATCAGGAGATGCTGAAGCTTCCCTATGTCACTAACAGATTCAGGATTTTCAGAATTCC
TTATATAGCTGCATGCAAGACAACTCATGAGTTATTGCATGGAGCTGGACAGATTAAAC
TCGTCCAACCAACCAAGAATTTGTAAAGGTTTATAAATCTGGTTTCGGTTGGGCGTTCA
TTAGACATCTCCGATTGAGCAGCTACCACGAGCTGCGAGAAGAGTTAGGGAAGATGTTT
GCTATCGAAGGGTTGTTGGAAGACCCCTTAGATCAGGCTGGCAGCTTGTATTCTGTTGAC
AAGGAAAATGATATTCTTCTCCTTGGTGATGACCCATGGGAGTCATTTGTGAATAACGTT

TGGTACATAAAGATACTATCACCAGAAGATGTGCATCAAATGGGAGATCATGGAGAAGGC
AGTGGTGGGTTATTCCCGCAAAACCCGACCCATCTCTAGAAGCTGCTTCGGTGTAGTCT
CATCATGCTACAACGCGGGAGCCCTTTGTTTCCCATTTGAAGTCGTTTCCACTCATCTTT
ATATGCCATTCTGTCGCATCTCTCTCGTTTGGACGTTTTAGAAAGAAACATAATCATAT
TTGTGAGTATGGGTCTGAAACTTTAGGACGTACTTTAGCTTGTATTAGACAGACACTCT
CGTCATAAACATAAGAACCTTTATGTAGCTGTCTCAGGGTAACTAACTTTTCTAG
>G1451 Amino Acid Sequence (domain in AA coordinates: 22-357)
MKLSTSGLGQQGHGEKCLNSELWHACAGPLVSLPSSGSRVYFPQGHSEQVAATTNKEV
DGHIPNYPSPQLICQLHNVTMHADVETDEVYAQMTLQPLTPPEQKETFVPIELGIPSK
QPSNYFCKTLTASDTSTHGGFSVPRRAAEKVFPPLDYTLQPPAQELIARDLHDVEWKFRH
IFRGQPKRHLTTGWSVVFSAKRLVAGDSVIFIRNEKNQLFLGIRHATRPQTIVPSSVLS
SDSMHIGLLAAAHASATNSCFTVFFHPRASQSEFVIQLSKYIKAVFHTRISVGMFRML
FETEESVRRYMGITITGISDLSVRWPNHRSVKVGWDESTAGERQPRVSLWEIEPLTT
FPMYPSLFLRLKRPWHAGTSSLPDGRDLGSLTWRGSGGEQQGLPLNYPVGLFPW
MQQRDLDSQMGTDDNNQQYQAMLAAGLQNIIGGDPLRQQFVQLQEPHHQYLOQSASHNSDL
MLQQQQQQQASRHLMAHTQIMSENLPQNMROEVSNOQAGQQQQQLQPDQNAYLNAFKM
QNGHLQQWQQQSEMPSPSFMKSDFTDSSNKFATTASPASGDNLLNFSITGQSVLPEQLT
TEGWSPKASNTFSEPLSLPQAYPGKSLALEPGNPQNPSLFGVDPDSGLFLPSTVPRFASS
SGDAEASPMSLTDSGFQNSLYSCMQDTTHELLHGAGQINSSNQTKNFVKVYKSGSVGRSL
DISRFSSYHELREELGKMFIEGLELPLRSGWQLVFVDKENDILLGLDDPWESFVNNVW
YIKILSPEDVHQMGDHGEGSGGLFPQNPTHL*
>G1478 (1..354)
ATGTGTAGAGGGTTTGAGAAAGAAGAGAGAAGAAGCGACAATGGAGGATGCCAAAGA
CTATGCACGGAGAGTCACAAAGCTCCGGTAAGCTGTGAGCTTTGCGGCGAGAACGCCACC
GTGTATTGTGAGGCAGACGCAGCTTTCTTTGTAGGAAATGCGATCGATGGGTCCATTCT
GCTAATTTTCTAGCTCGGAGACATCTCCGGCGCGTGATCTGCACGACCTGTGCGAAGCTA
ACTCGTCGATGTCTTGTGCGGTGATAATTTTAATGTTGTTTTACCGGAGATAAGGATGATA
GCAAGGATTGAAGACATAGTAGTGATCACAAATTCCTTTGTGTTTCTCTGA
>G1478 Amino Acid Sequence (domain in aa coordinates: 32-76)
MCRGFEKEEERRSDNGCQRLCTESHKAPVSCELCGENATVYCEADAFLCRKCDRWVHS
ANFLARRHLRRVICITCRKLTRRCLVGDNFNVVLPEIRMIARIEEHSSDHKIPFVFL*
>G1496 (116..1123)
AAACCCACCAAATAACTCAGAGCTTTTTTGCATTTTTTCCCATTTCTATTTTGT
ACTTTTGGTCTCACTTTAAAGATCATAAGTTGAAAGATTTCTGCAGAGAACAATATGTT
GGAAGGTCTTGTCTCTCAAGAAAGCTTGTCTTAAACTCTATGGACATGTCTGTACTTGA
AAGGCTTAAATGGGTACAACAGCAACAACAGCAACTGCAACAAGTTGTGTCCCATAGCAG
TAATAATTCACCTGAACCTTCTCAGATACTTCAGTTCCATGGAAGCAACAATGATGAGTT
GTTGGAGAGTAGTTTTCAGCCAAATTTCAAATGCTTGGATCTGGTTTTGGACCAAACATATAA
CATGGGTTTTTGGTCTCCACATGAATCCATTTCAAGAACAAGTAGCTGCCATATGGAACC
TGTGGATACAATGGAGGTTTTGTTGAAGACCGGTGAAGAAACCAGAGCCGTTGCCTTGAA
GAACAAGAGAAAACCAGAGGTTAAGACAAGGAAGAGCAAAAGACAGAGAAGAAGATCAA
AGTAGAGGCTGAGACAGAGTCAAGCATGAAAGGAAAATCAAACATGGGAAACACTGAAGC
ATCTTCAGACACTTCAAAGGAGACATCGAAAGGAGCTTCAGAGAATCAGAAATTAGATTA
TATCCACGTGAGAGCTCGTCGAGGCCAAGCCACTGACAGACACAGCTTAGCAGAAAGGGC
GAGAAGAGAAAAGATCAGCAAGAAAATGAAATATCTGCAAGATATTGTGCCTGGATGCAA
TAAGGTCACAGGAAAAGCTGGTATGCTTGATGAGATCATCAATTATGTTCAATGTCTCCA
AAGACAAGTCAAGTCTCTGTGTCGATGAAACTTGTCTTGAACCGGAACCTAGAGCTTGC
CGTGAAGATGTATCCGTAAAACAGGCTTACTTTTCAAATGTAGTTGCTTCAAAGCAATC
AATAATGGTTGATGTGCCATTGTTTCCGTTAGACCAGCAAGGATCTCTAGATTGTCTGC
GATAAACCCGAACCAAACGACATCTATCGAAGCTCCATCTGGAAGCTGGGAAACCTAATC
ACAGAGTCTCTACAACACATCTAGCCTCGGTTTTTCACTACTAAGCAAGATTCAATTGAAAC
AACATGGTTGACATCAATCAATCATCAAAATCAGAAGCAAATTCATTACATTTGCTCAT
CAAAGTAGTAATTTGAAATTTGGTTAATGCATTATCCTTTGATCCTTGTCTTCTGATAT
TAAACCAGAAAGAACTGGAGATAGCAATCCAATGATCTTGTACCA
>G1496 Amino Acid Sequence (domain in AA coordinates: 184-248)
MLEGLVSQESLSLNSMDSVLERLKWVQQQQQLQVVS HSSNNSPELLQILQFHGSNND
ELLESSFQFQMLGSGFGPNYNMFGPPHESISRTSSCHMEPVDTMEVLLKTGEETRAVA

LKNKRKPEVKTREEQKTEKKIKVEAETESSMKGKSNMGNTASSDTSKETSKGASENQKL
DYIHVRARRGQATDRHSLAERARREKISKMKYLQDIVPGCNKVTGKAGMLDEIINYVQC
LQRQVEFLSMKLAVLNPELELAVEDVSVKQAYFTNVVASKQSIMVDVPLFPLDQQGSDDL
SAINPNQTTTSIEAPSGSWETQSQSLYNTSSLGFHY*

>G1526 (1..3090)

ATGGGAACGAAAGTCTCAGACGATCTTGTTTCCACCGTCAGATCAGTCGTGGGTTCGGAT
TACTCAGATATGGATATAATCAGGGCTTTACACATGGCGAATCATGATCCAACGGCTGCT
ATCAATATAATCTTCGACACTCCAAGTTTCGCCAAACCTGATGTAGCCACTCTACCCCG
AGCGGCTCTAATGGAGGGAAGCGAGTTGATAGTGGATTAAAGGGCTGTACTTTTGGTGAC
AGCGGAAGTGTGGAGCGAATCATCGCGTGGAGGAAGAAAATGAGAGTGTAAATGGTGGA
GGAGAAGAGAGTGTTCAGGGAATGAGTGGTGGTTTGTGGTTGTTCTGAATTGGCTGGG
TTATCGACATGTAAAGGAAGGAAATTGAAGTCTGGTGATGAATTGGTGTTCACGTTTCCG
CATAGTAAAGGATTAAAGCCTGAGACTACGCCTGGGAAGCGCGGTTTGGGCGGGGAAGG
CCAGCTTTGCGTGGTGCTTCTGATATCGTTAGGTTCTCTACAAAGGATTTCAGGAGAGATT
GGTAGAATACCAAACGAGTGGGCTCGGTGCTCTTCTACCACTTGTGAGAGACAAGAAAATT
AGGATAGAAGGCAGTTGCAAGTCGGCGCCTGAAGCTTTGAGCATCATGGATACAATTCTT
CTGTCTGTAAAGCGTGTACATTAATAGTTCCATGTTTCAAAGCATAGTGGACTTCATTT
AAGACAGCTAGTAATACGGCAGAGGAATCAATGTTCCATCCTCTCCCAAATCTCTTCGG
TTACTCGGTTTGATCCCTTTAAGAAGGCAGAGTTTACTCCAGAGGATTTTACTCTAAG
AAGCGACCTTTGAGTTCCAAGGATGGTTCTGCTATTCTTCTTCTGCTTCAATTAAAC
AAGGTCAAGAATATGAATCAAGATGCAAACGGAGATGAAAATGAGCAGTGTATCAGCGAT
GGTGATCTTGATAACATTGTTGGTGTGGGGACAGTTCGGATTAAAGGAAATGGAACT
CCACATACACTTCTGTGTGAGCTTCGTCCATACCAAAGCAGGCACCTTCATTGGATGACC
CAACTGGAGAAAGGAAATTGCACATGATGAGGCAGCAACAATGCTTCACCCGTGTTGGGAA
GCATACTGTTTAGCAGACAAGAGGGAAGTGGTTGTCTACCTGAATTCTTTTACTGGTGAT
GCTACAATACACTTCCCTAGCACACTTCAAATGGCAAGAGGAGGAATATTAGCAGACGCA
ATGGGTCTTGGAAGAGCTGTAATGACCATATCCCTTTTGCTTGCCCATTCCTGGAAAGCT
GCATCAACTGGGTTTCTATGCCCCAACTATGAAGGAGACAAAGTGATCAGCAGTTCTGTA
GATGATCTCACTAGTCCCCCGGTGAAGGCAACCAAATTTCTAGGCTTTGATAAGAGGCTT
CTTGAACAAAAAGTGTACTTCAAATGGTGGTAACCTGATTGTATGTCCGATGACACTT
TTAGGACAGTGGGAAGACAGAGATTGAAATGCATGCAAAGCCTGGGTCTCTATCTGTCTAT
GTTCACTATGGGCAAAGCAGGCCGAAGGATGCAAACCTCTTTCCAGAGTGATGGTGA
ATCACCACATATGGAGTTCTAACATCCGAATTCTCGCAAGAGAACTCAGCAGACCATGAA
GGAAATTTATGCAGTTTCGATGGTTTAGGATTGTTCTTGACGAGGCACATACCATCAAAAC
TCAAAAAGCCAAATTTCTTGGCTGCTGCAGCTCTGGTTGCTGATAGGCGTTGGTGTCTT
ACGGGTACTCTTATTCAGAACAACTCTGGAGGATTATACAGCCTTCTACGGTTTTTGAGG
ATTGAACCATGGGGAACCTGGGCATGGTGGAAATAAAGTTGTCCAAAAGCCATTTGAAGAG
GGTGATGAGAGAGGGTTAAAGCTAGTGCAGTCTATCTTAAAACCTATCATGCTTAGGAGA
ACAAAGTCTAGCACAGACCGAGAAGGAAGGCCGATTCTTGTTCTACCCCTGCTGATGCA
CGGGTCACTTACTGTGAACCTTTCGGAGTCTGAGAGGGATTCTACGACGCGCTATTTAAA
AGATCCAAGGTCAAATTTGATCAATTTGTTGAACAAGGCAAAGTTCTTCATAACTATGCT
TCGATCCTGGAACCTGCTTTGCGTCTTCGACAATGTTGTGATCACCCTTTTGTAGTAATG
AGTCGAGGGGATACAGCGGAATACTCTGATCTGAATAAGCTTTCTAAACGTTTCCTTAGT
GGAAAGTCTTCTGGCTTAGAAAAGGGAAGGAAAAGATGTACCGTCAGAGGCTTTTGTTCAG
GAGGTGGTAGAGGAACCTGCGCAAAGGAGAGCAAGGAGAGTGTCCAATATGCCTTGAAGCA
CTTGAGGATGCTGTATTAAACGCCATGTGCTCATAGATTATGTGCTGAGTGTCTCTTGGCA
AGTTGGAGAAATTTCTACTTCTGGGTTATGTCTGTGTAGGAACACTGTAAGCAAACAA
GAACATCATCAGCACCAACCGAAAGTAGATTCCAGGTTGACGTGGAAAAGAAATGGGTG
GAATCATCGAAAATCACTGCTCTTCTGGAAGAGCTTGAAGGTCTTCGTTCTTCAGGCTCT
AAGAGCATTCTCTTTAGCCGATGGACCGCTTTCTCGATCTCTCTCAAATTTCCCTCTCT
CGGAATAACTTTTTCATTGTTCCGCTTTGATGGCACGCTAAGTCAGCAGCAACGAGAGAAG
GTCTTTAAAGAATTTTCCGAAGATGGCAGTATCCTGGTACTGTTGATGTCTCTAAAAGCT
GGTGGCGTTGGGATAAATCTAACAGCTGCGTCCAATGCTTTGTGATGGATCCATGGTGG
AACCAGCGGTAGAGGAACAAGCTGTTATGCGTATTTCATCGTATAGGGCAAACCTAAGGAA
GTCAAATCAGAAGATTTCATCGTTAAGGGAACGGTTGAAGAGAGAATGGAGGCGGTTTCAG
GCGAGGAAGCAGAGAATGATCTCTGGGGCTTTAACCGATCAAGAAGTACGAAGTGCACGT
ATAGAGGAACCTCAAGATGTTATTTACCTGA

>G1526 Amino Acid Sequence (domain in AA coordinates: 493-620, 864-1006)

MGTKVSDDLVSTVRSVVGSDYSDMDIIRALHMANHDPTAAINIIFDTPSFAKPDVATPTP
SGSNGGKRVDSGLKGCTFGDSGSGVGNHRVEEENESVNGGGEESVSGNEWWFVGCSELG
LSTCKGRKLKSGDELVFTFPHSKGLKPETTPGKRGFGRGRPALRGASDIVRFSTKDSGEI
GRIPNEWARCLLPVLRDKKIRIEGSCSAPEALSIMDTILLSVSVYINSSMFQKHSATSF
KTASNTAEESMFHPLPNLFRLLGLIPFKAEFTPEDFYSKKRPLSSKDGSAPIPTSLQLN
KVKNMNQDANGDENEQCISDGLDNIVGVGDSSGLKEMETPHTLLCELRPYQKQALHWM
QLEKGNCTDEAATMLHPCWEAYCLADKRELVVYLNSFTGDATIHFPSTLQMGAGGILADA
MGLGKTVMTISLLLAHSWKAASGFLCPNYEGDKVISSSVDDLTPSPVKATKFLGFDKRL
LEQKSVLQNGGNLIVCPMTLLGQWKTEIEMHAKPGSLSVYVHYGQSRPKDAKLSSQSDVV
ITTYGVLTSEFSQENSADHEGIYAVRWFRIVLDEAHTIKNSKSQISLAAAALVADRRWCL
TGTPIQNLEDLYSLRFLRIEPWGTWAWWNKLQKPFEEGDERGLKLVQSILKPIMLRR
TKSSTDREGRPIVLPPADARVIYCELSERDFYDALFKRSKVKFDQFVEQGVKLVHNYA
SILELLRLRQCCDHPFLVMSRGDTAEYSDLNKLSKRFLSGKSSGLEREGKDVPEAFVQ
EVVEELRKGEQGECPICLEALEDVLTCAHRLCRECLLASWRNSTSGLCPVCRNTVSKQ
ELITAPTESRFQVDVEKNWVESKITALLLEEGLRSGSKSILFSQWTAFLDLLQIPLS
RNNFSFVRLDGTLSQQQREKVLKEFSEDGSILVLLMSLKAGGVGINLTAASNAFVMDPWW
NPAVEEQAVMRIHRIGQTKVKIRRFIVKGTVEERMEAVQARKQRMISGALTDQEVRSAR
IEELKMLFT*

>G1543 (1..828)

ATGATAAACTACTATTTACGTACATATGCACATACACATATAAACTATATGCTCTATAT
CATATGGATTACGCATGCGTGTGTATGTATAAAATATAAAGGCATCGTCACGCTTCAAGTT
TGCTCTCTTTTATATTAACCTGAGAGTTTTCTCTCAAACCTTTACCTTTTCTTCTCGATC
CTAGCTCTTAAGAACCCTAATAATTCATTGATCAAAATAATGGCGATTTTGCCGGAAC
TCTTCAAACCTTGGATCTTACTATCTCCGTTCCAGGCTTCTCTTCATCCCTCTCTCCGAT
GAAGGAAGTGGCGGAGGAAGAGACCAGCTAAGGCTAGACATGAATCGGTTACCGTCGTCT
GAAGACGGAGACGATGAAGAATTCAGTCACGATGATGGCTCTGCTCTCCGCGAAAGAAA
CTCCGCTCTAACCAGAGAACAGTCACGCTCTTCTGAAGATAGTTTCAGACAGAATCATACC
CTTAATCCCAAACAAAAGGAAGTACTTGCCAAGCATTTGATGCTACGGCCAAGACAAATT
GAAGTTTGGTTTCAAACCGTAGAGCAAGGAGCAAAATTGAAGCAAAACCGAGATGGAATGC
GAGTATCTCAAAGGTGGTTTGGTTTCATTAACGGAAGAAAACACAGGCTCCATAGAGAA
GTAGAAGAGCTTAGAGCCATAAAGGTTGGCCCAACAACGGTGAACCTCTGCCTCGAGCCTT
ACTATGTGTCTCGCTGCGAGCGAGTTACCCCTGCGCGAGCCCTTCGAGGGCGGTGGTG
CCGTTTCCGGCTAAGAAAACGTTTCCGCGCAAGAGCGTGATCGTTGA

>G1543 Amino Acid Sequence (domain in AA coordinates: 135-195)

MIKLLFTYICTYTYKLYALYHMDYACVCMYKYKGIIVTLQVCLFYIKLRVFLSNFTFSSSI
LALKNPNNSLIKIMAILPENSSNLDLTISVPGFSSSPLSDEGSGGGRDQLRLDMNRLPSS
EDGDDEEFSHDDGSAPPRKKLRLTREQSRLLDSFRQNHTLNPQKEVLAKHMLRPRQI
EVWFQNRARRSKLKQTEMECEYLKRWFGSLTEENHRLHREVEELRAIKVGPTTVNSASSL
TMCPRCERVTPAASPSRAVVPVPAKKTFFPPQERDR*

>G162 (101..619)

AGACATACAACACCAAATCTTCTTCTTCCACCAACATATTCACCTTTCACAGCAAAAAA
ACGAGAGGTTCTCTCTTATTCGTACCGTTTTAGCAAACAAATGGGTCCGAGAAAGATCAA
GATGGAGATGGTTCAAGGACATGAACACACGACAGGTTACCTTTTCAAACGGAGGACTGG
TTTGTTCAGAAGGCGAGCGAGTTAGCCACGCTCTGCAACGCTGAGTTGGGCATCGTTGT
CTTTTCCACAGGAGGCAAGCCTTCTCCTACGGGAAACCGAATCTTGATTCTGTTGCAGA
GCGATTATGAGAGAATATGATGATTGAGACAGTGGCGATGAAGAAAAAAGTGGTAATTA
CAGGCTTAACTGAAGAGGCTGAGTGAACGCTCTCGATTGCTCAACCAAGAGGTTGAAGC
TGAGAAGGAACGAGGCGAGAAGAGTCAGGAGAAGCTTGAATCTGCTGGGGATGAGAGATT
CAAGGAGTCCATTGAGACGCTTACCCTCGATGAACCTCAATGAATACAAAGATAGGCTTCA
GACAGTCCATGGTAGGATTGAAGGTCAAGTCAATCACTGACAGGCTTCGTCTTGCCTCAT
GCTTCTCTCCAGAAAATAGCTAGACCGACTTGTAGAGTTACATTCTATTTTTGTATCA
GCCTACAGAACTTACCAACACATGAAAGTTATTGCTGGTGTAGAATTTCTGTCTATCTAT
GGGGTGTGACTTTCTATTGACATCAAATGAAAATGTACCTGGAAATTTGTCTGTATTAA
TCTCAAGTGTACTTGTCTAAACTTGATCAGCTTTTTTCGCAAAAAA

>G162 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRRIKMEMVQDMNTRQVTFSKRRTGLFKKASELATLCNAELGIVVFSPPGKPFYSYKPK

NLDSVAERFMREYDDSDSGDEEKSGNYRPKLKRLSERLDLLNQEVEAEKERGEKSQEKLE
SAGDERFKESIETLTLDLNEYKDRLQTVHGRIEGQVNLQASSCLMLLSRK*

>G1640 (168..1196)

TTCCGCAGATCCTTCCTCTATATAAGGAAGTTCATTTTTCATTTGGAGAGGTTTCGCTGACA
AGCTGCTCTAGCTTATCTGGTACCGTCGACCTCTCACTCAAGGGTCCAAAAGTGTCTTCT
CTTTTTTCAGTTTCTCTTTCTCTTTTGGACAGAAGAGACCGAGAAGCAATGGGAAGGGCTC
CGTGTGTGAGAAAATCGGGTTGAAGAGAGGGAGATGGACAGCCGAGGAAGATGAGATCC
TCACCAAGTATATTCAGACCAATGGTGAAGGTTCTTGGCGATCTTTGCCTAAGAAAGCTG
GATTGTTGAGATGTGGAAGAGCTGTAGACTAAGGTGGATAAACTACTTAAGAAGAGACT
TAAAAAGAGGAAATATTACTTCCGACGAAGAAGAAATAATCGTCAAGTTGCATTCCCTTC
TCGGCAACAGATGGTCACTTATTGCAACACATCTACCAGGAAGAACAGACAACGAAATTA
AAAATATTGGAACACATCTCAGCCGCAAAATCTATGCCTTCACTGCCGTTTCCGGAG
ATGGACACAATCTACTCGTCAACGATGTAGTCTTGAAGAAATCTTGTTCATCGTCTTCTG
GAGCCAAGAACAATAACAAGACCAAGAAGAAGAAGGGAAGGACTAGTAGGTCAATCCA
TGAAGAAACACAAGCAAAATGGTGACGGCCTCACAATGTTTCTCACAACCTAAGGAGCTAG
AGAGTGATTTTCACTGAGGGAGGGCAAAATGGTAATTTTGAAGGAGAGTCTTTGGGGCCTT
ATGAGTGGTTGGATGGTGAAGTGAACGGCTCTTGAGTAGTTGTGTCTGGGAATGCCTA
GTGAAGAGGCTGTGATTTGGAGTAAATGATGAAAAGGTGTGTGAGAGTGGGGACAATAGTA
GTTGTTGTGTTAATTTGTTTGAAGAAGAACAAGGAAGCGAGACAAAGATTGGTCACGTAG
GAATCACAGAGGTTGATCATGATATGACGGTGGAAAGAGAAAGAGAGGGAAGTTTTTTAA
GTTCAATTCAAATGAAAATAATGATAAAGATTGGTGGGTTGGTCTATGTAATCTTTCAG
AAGTTGGGTTTGGGGTTGATGAGGAGTTGCTTGATTGGGAGTTTCAAGGTAATGTCACTT
GTCAAAGTGATGATCTATGGGATCTCTCAGATATTGGAGAGATAACATTGGAGTGATTGT
ACCGAGCAAGTGGATTGGCGGCCGCTCTAGACAGGCCTCGTACCGGATCTCTAGCTAGAG
CTTTCGTTTCGTATCATCGGTTTCGACAACGTTTCGTCAAGT

>G1640 Amino Acid Sequence (domain in AA coordinates: 14-115)

MGRAPCEKIGLKRGRWTAEEDEILTKYIQTNEGSWRSLPKKAGLLRCGKSCRLRWINY
LRDLKRGNITSDEEEIIVKLHSLGNRWSLIATHLPGRDNEIKNYWNSHLRKYAFT
AVSGDGHNLLVNDVVLKSCSSSSGAKNNNKTKKKKKGRTRSRSMSKKHKQMVTAQCFSQ
PKELESDFSEGGQNGNFEGESLGPYEWLDGELERLLSSCVWECTSEEAVIGVNDEKVCES
GDNSSCCVNLFEEEQGSETKIGHVGITEVDHDMTVEREREFSFLSSNSNENNDKDWVWGL
CNSSEVGFVDEELLDWEFQGNVTCQSDDLWDLSDIGEITL*

>G1644 (1..348)

ATGAAATTGATTGGAAAGACTGTGCTTTGATGACTTACACCGAACTCATTTTGGGT
TTCTGCAATGTTTTAATGTTGATCTGCAGGAGGACTAGTGGACCTATGAGACGAGCAAAA
GGTGGTTGGACTCCAGAGGAGGATGAGACACTTAGACGAGCAGTTGAAAAGTATAAGGGG
AAGAGGTGGAAGAAAATAGCGGAATTTTTCCAGAGAGAACACAAGTCCAATGCTTGCAC
AGGTGGCAGAAAGTTCTTAATCCAGAGCTTGTTAAAGGACCTTGGACTCAAGAGGTTCTC
TTATCATTTTCATGTTCTGAAACTTTTTTGGTTTTCATTTTACGTAA

>G1644 Amino Acid Sequence (conserved domain in AA coordinates: 39-102)

MKLIDWKDCALMTYTELILGFCNVLMILICRRTSGPMRRAKGGWTPPEEDTLRRAVEKYKG
KRWKIAEFFPERTQVQCLHRWQKVLNPELVKGPWTQEVLLSFSCSETFFGFHFT*

>G1646 (34..786)

GATCTTTTGATCCAATCACAAGGCAAGATCCAATGGACAATAACAACAACAACAAC
CAGCAACCACCACCAACCTCCGCTATCCACCTGGCTCCGCCGTCACAACCGTAATCCCT
CCTCCACCATCTGGATCTGCATCAATAGTCACCGGAGGAGGAGCGACATACCACCACCTC
CTCCAGCAACAACAGCAACAGCTTCAAATGTTCTGGACATACCAGAGACAAGAGATCGAA
CAGGTAAACGATTTCAAAAACCATCAGCTCCCTCTAGCTCGTATCAAAAAATCATGAAA
GCTGATGAAGATGTGCGTATGATCTCCGCCGAAGCACCGATTCTCTTCGCGAAAGCTTGT
GAGCTTTTCATTCTCGAACTTACGATTAGATCTTGGCTTCACGCTGAAGAGAACAACGCT
CGTACGCTTCAGAAAAACGATATCGCTGCTGCGATTACTAGAACCGATATCTTCGATTTC
CTTGTGATATTGTTCTTAGGGAAGAGATCAAGGAAGAGGAAGATGCAGCATCGGCTCTT
GGTGGAGGAGGTATGGTTGCTCCCGCCGCGAGCGGTGTTCTTATTATTATCCACCGATG
GGACAACCGGCGGTTCTGAGGGGATGATGATTGGAAGACCGGCGATGGATCCTAGCGGT
GTTTATGCTCAGCCTCCTTCTCAGGCATGGCAAAGCGTTTGGCAGAATTCAGCTGGTGGT
GGTGATGATGTGCTTATGGAAGTGAGGAAGTAGCGCCATGGTAATCTCGATAGCCAA
GGGTAAGTGAATTCTAGTAG

>G1646 Amino Acid Sequence (domain in AA coordinates: 72-162)
MDNNNNNNNQPPPTSVYPPGSAVTTVIPPSPSGSASIVTGGGATYHLLQQQQQLQMF
WTYQRQEIEQVNDFKNHQLPLARIKKIMKADEDVRMISAEAPILFAKACELFILELTIRS
WLHAEENKRRTLQKNDIAAAITRTDIFDFLVDIVPREEIKEEEDAASALGGGGMVAPAAS
GVPIYYPPMGQPAVPGMMIGRPAMDPSGVYAQPPSQAWQSVWQNSAGGGDDVSYSGSGS
SGHGNLDSQG*

>G1672 (239..1399)
CCATTCTGACGTCCGGGATGACGCACAATCCCACTATCCTTCGCAAGACCCTTCCCTCTA
TATAAGGAAGTTTCATTTTCATTTGGAGAGGACACGCTGACAAGCTGACTCTAGCAGATCTG
GTACCGATCACTCCCGTCTTTATCAAATTCCTTCTTCTTCTTACATTTTCCCTATCCAATC
GATCTCACGCAGATCTGATCAATTTCTCATCAAATCATTTAGAGATCAAAAGAAAACCTAT
GAAGAATAGTAAATGTAACCTCATAGATTCAAAGCTCGAAGAACATCATCATCTTTGCGG
ATCAAAACATTGTCTGGATGTGGTCGCATGATTCAAGCTGCTACTAAACCAAATTGGGT
TGGATTGCCGGCAGGAGTGAAATTCGATCCGACAGATCAAGAACTTATAGAACATTTAGA
AGCAAAAGTGAAGGGAAAAGAAAATAAGAAATGGTCGTCGTCTCATCCACTTATAGA
TGAATTTATTCCACCATTGATGGAGAAGATGGAATATGTTACACTCATCTCAGAAGCT
TCCAGGGGTGACAAGAGATGGCTTGAGCAAACACTTCTTCCACAAACCATCAAGAGCTTA
CACAAACCGGAACAAGAAAACGACGTAAATAATTCAAACCGATCACGACTCTGAGTTAAC
CGGATCATCAGAAACCGGTGGGCACAAAACGGGGCAAAACAAGACCGGTTATGATCAACGG
TCAACAAAGAGGATGCAAGAAGATATTAGTACTCTACACAAACTTCGGCAAGAATCGTCG
ACCGGAGAAAACAAATTGGGTGATGCATCAATATCATTAGGGATTAATGAGGAAGAGAG
AGAAGGAGAACTTGTGGTCTCCAAGATATTTTATCAGACACAACCAAGACAGTGTGTAG
TAATACTAATTGGTCTGATCACCATGGTTCCAAGGACGTGATCGGAATTGGTGTGCGAGA
TGAGATTTCAGCGTAGCTGCCACGTTGCAGAGTCTTGGCTCCGGTGACGTCGTTTCTAG
GGTTAATATGCATCCCCATACAAGATCCTTTGATGAGGGGACAGCCGAAGCTTCAAAGGG
AAGAGAGAACCAGCATGTGTCTGGCACGTGCGAGGAAGTACATGATGGGATCATAACATC
ATCAATGTCTATCATATGATTCATGATCATCATAATCAACATCATCAAATCGGAGA
TAGAAGAGAATTTACATGTCTATCATATCCCATGACCCCTACTATCACATCAACA
TGAGTCAATCTTCCATGTTACAAGTACTATGCCCTTTTCAGCGGCAGCAATTAAGGGGTG
GTCGTCTGGTTCCGGGATTAGAAGACCTAATTATGGGTTGTACCACAGCTACGTGTACAGA
AGACAATAATCACAATGATTAAATTCGCAGGAGCATTCAAGAACAAACCCCTCAGCGAAA
TGCAGAGTGGTTAACGTTTCCACAATTTCTGGAACCAAGCCGAATCAGATGATCAAAACCG
AAGATTTTAAACAGAACCAAAAGGAAGCAGAGAAATCTTGCAAAAAGCTCCTGCTTAGCTG
TTGATCAATGCCGAAATGCTGAGCTATGACTGACTAGTCTCTGCCATTTAACTTACAAT
ATCACCAGAGGTTGCGATGAATGTTGATTCGCTCAAAGGAGAGCGGCCGCTCTAGACAGG
CCTCGTACCG

>G1672 Amino Acid Sequence (conserved domain in AA coordinates: 41-194)
MKNSKCNLIDSKLEEHHHLGSKHCPGCGRMIQAATKPNWVGLPAGVKFDPTDQELIEHL
EAKVKGEENKKWSSSHPLIDEFIPTIDGEDGICYTHPQKLPGVTRDGLSKHFFHKPSRA
YTTGTRKRRKIQTDDHSELTEGSETRWHKTGKTRPVMINGQQRGCKKILVLYTNFGKNR
RPEKTNWVMHQYHLGINEEEREGLVVS KIFYQTQPRQCVSNNTNWS DHHGS KDVI GIGVG
DEISSVAATLQSLGSGDVSVRVNMHPHTRSFDEGTAEASKGRENQHVSGTCEEVHDGIIT
SSMSSHMIHDHNQH HQIGDRREFHMS SYPMTPTITSQHESI FHV TSTMPFQRQQLRG
RSSGSGLEDLIMGCTTATCTEDNNHK*

>G1677 (24..1037)
CAGTACTAATTCTGTGTGTGTTAATGGTTCTAGTTATGGATGATGAAGAGAGTAACAACG
TTGAAAGATATGACGACGTGCTATTGCCAGGGTTTAGGTTCCATCCCACTGATGAAGAAC
TCGTAAGTTTCTACTTGAAACGGAAGGTTTTACACAAATCTCTTCCCTTTGATCTCATCA
AGAAAGTCGACATTTACAAATACGATCCATGGGACCTCCCAAAGCTTGCGAGCGATGGGGG
AAAAAGAGTGGTACTTTTATTGTCTTAGAGACAGGAAATACCGCAACAGCACAAAGACCTA
ACCGAGTAAC TGGAGGTGGCTTCTGGAAAGCAACCGGAACAGACCGGCTATATACTCAT
TGGACTCCACTCGATGCATCGGTTTGAAGAAATCACTTGTGTTCTACCGTGGTTCGAGCTG
CTAAAGGAGTCAAAACCGATTGGATGATGCATGAATTCGTCTCCCTTCTCTCTGACT
CTCATCACTCATCATATCCCAATTACAATAACAAGAGCAACACCTTAACAATAACAACA
ACAGCAAGGAGCTTCCTTCAAACGATGCTTGGGCGATATGTAGAATATTTAAGAAGACAA
ATGCAGTATCCTCACAAAGATCAATCCCACAATCTTGGGTTTATCCAACGATTCCTGACA
ACAATCAACAGTCACACAACAACACCGCAACTCTCTTAGCTTCATCAGACGTTCTCAGCC

ACATATCAACAAGACAAAACCTTTATTCTCTCTCCAGTCAACGAACCCGCAAGCTTCACAG
AATCAGCTGCTTCTTACTTCGCGTCTCAGATGCTCGGAGTCACGTACAATACAGCCAGAA
ACAACGGAACAGGGGATGCTCTGTTTCTGAGAAACAATGGAACAGGGGATGCTCTGGTTC
TGAGCAACAATGAGAATAACTACTTCAACAACCTTGACTGGAGGGTTGACTCATGAGGTTT
CGAATGTAAGATCAATGGTGATGGAGGAGACTACGGGGAGTGAGATGTCGGCGACGTCGT
ATTCCACTAACAATAAGATCATAGTACTATTAACTTGAATTAGTGTAGACGTTGATC
ATCGCTAATATGTATTAATTTTCTTGTCTTACTATAAACGAAAAA

>G1677 Amino Acid Sequence (conserved domain in AA coordinates: 17-181)

MVLVMDDEESNNVERYDDVLPGRFPHPTDEELVSFYLRKVLHKSPLPDLIKKVDIYKY
DPWDLPLKLAAMGEKEWYFYCPRDRKYRNSRPNRVTTGGGFWKATGTDRIYSLDSTRCIG
LKKS LVPYRGRAAKGVKTDWMMHEFRLPSLSDSHHSSYPNYYNNKKQHLNNNNNSKELPSN
DAWAICRIFKKTNAVSSQRSIPQSVVYPTIPDNNQQSHNNTATLLASSDVLISHISTRQNF
IPSPVNEPASFTESAASYFASQMLGVTYNTARNNGTGDALFLRNNGTGDALVLSNNENNY
FNNLTGGLTHEVPNVRSMVMEETTGESEMSATSYSTNN*

>G1765 (139..966)

TCCTTCGCAAGACCCCTCTCTATATAAGGAAGTTCATTTTATTGGAGAGGACACGCTG
ACAAGCTGACTCTAGCAGATCTGGTACCGTCGACAAGAATGACTTGATTGGTGTTCTAAA
GAGATCGATGTAGTGAAGATGAGTGGCGAAGGTAACCTTAGGTAAGGATCATGAAGAAGAA
AACGAAGCACCCTTCTTGGGTTTCAAGTTTTCATCCGACGGATGAAGAGCTTTTAGGATAC
TATCTTCGAAGAAAAGTAGAGAACAAACCATCAAACCTCGAATTATCAAACAGATCGAT
ATCTATAAGTACGATCCTTGGGATCTTCCAAGAGTGAGCAGCGTCGGAGAAAAGGAGTGG
TACTTCTTCTGCATGAGAGGTAGGAAATACAGGAATAGCGTTTCGACCAAACCGAGTGACC
GGTTCAGGTTTCTGGAAAGCCACTGGTATTGATAAACCGGTTTACTCCAATCTTGACTGT
GTTGGTCTCAAGAAATCTCTGGTTTACTATCTTGGTTTCAAGCGGTAAGGCACCAAAC
GATTGGATGATGCATGAATTCCGCCTCCCTCCACCACGAAAACCGACTCTCCAGCTCAA
CAAGCAGAGGTATGGACACTTTCGAGAATCTTCAAACGAGTCACATCTCAAAGAAACCCA
ACCATCTTACCACCAAACCGAAAACCGGTTATCACTTTAACCGACACTTGTCTAAGACC
AGCAGCTTAGATTCCGACCACAGGACCCGTACAGTAGATTCCATGTCCCACGAGCCG
CCGCTTCCACAGCCACAGAAATCCTTATTGGAACCAACATATAGTTGGTTTAAATCAACCG
ACATATACTGGTAATGATAATAACCTCCTGATGAGTTTCTGGAACGGCAACGGTGGAGAT
TTCATAGGAGACTCAGCAAGTTGGGATGAACCTAGATCTGTTATAGATGGCAACACTAAA
CCCTAGTAATAAAGTTTCTTTTTCAGCTTTGTACAAAAGATAAAACAAACGGCAACC
GCTCTAGACAGGCCTCTGACCGGATCCTCTAGCTAGAGCTTTCGTTTCGTATCATCGGT
TTCGACAACGTTTCGT

>G1765 Amino Acid Sequence (conserved domain in AA coordinates: 20-140)

MSGEGNLGKDHEEENEAPLPGRFPHPTDEELLYLRRKVENKTIKLELIKQIDIYKYDP
WDLPRVSSVGEKEWYFFCMRGRKYRNSVRPNRVTTGGGFWKATGIDKPVYSNLDVGLKKS
LVYYLGSAGKGTCTDMMHEFRLPSTTKTDSPAQQAQAEVWTLCRIFKRVTQSRNPITLPPN
RKPVITLTDTCSTSSLDSDHTSHRTVDSMSHEPPLPQPQNPYWNQHVGFNQPTTGTND
NNLLMSFWNGNGGDFIGDSASWDELRSVIDGNTKP*

>G1777 (97..1878)

CTCGTACTTTATCACCTCCGTCGTTCTATAATACTCTCTTCCGTCAATCATATCATTTGT
CGACAATTTTCTTCTGATCAGTTTAAAAATTGATCCATGGATGATAATTTAAGCGGCGAG
GAAGAAGATTACTATTACTCCTCCGATCAGGAATCTCTCAACGGGATTGATAATGATGAA
TCCGTTTCGATACCTGTTTCTTCCCGATCAAATACTGTCAAGGTTATTACGAAGGAATCA
CTTTTGGCTGCACAGAGGGAGGATTTGCGGAGAGTGATGGAATTGTTATCGGTTAAGGAG
CACCATGCTCGGACTCTTCTTATACATTACCGATGGGATGTGGAGAAGTTGTTTGCTGTT
CTTGTGAGAAAGGSAAGATAGCTTGTTTCTGGTGCTGGTGTTACACTTCTTGAAAAC
CAAAGTTGTGATTCTTCCGTTTCTGGTTCTTCTCGATGATGAGTTGTGATATCTGCGTA
GAGGATGTACCGGTTTACAGCTGACAAGGATGGAGTGTGGCCATAGCTTTTGCAATAAC
TGTTGGACTGGGCATTTTACTGTAAAGATAAATGAAGGTCAGAGCAAAGGATTATATGC
ATGGCTCATAAGTGTAATGCTATTTGTGATGAAGATGTTGTCAGGGCTCTAGTTAGTAAA
AGCCAACAGATTTAGCTGAGAAGTTTGATCGTTTTCTTCTTGAGTCGTATATCGAAGAT
AACAAAATGGTGAAGTGGTGCCGAGTACTCCTCATTGTGGGAATGCCATACGTGTTGAG
GATGACGAGCTCTGTGAGGTTGAATGCTCTTGTGGTTTGCAGTTCTGTTTCAGTTGTTCA
TCTCAAGCTCACTCCCTTGTCTTGTGTGATGTGGGAATATCGAGAAAGAAGTGCTTT
GATGAGTCCGAGACTGTTAATTGGATAACTGTTACACAAAGCCGTGTCCCAATGTCCAC

AAGCCTGTTGAAAAGAATGGTGGATGCAATCTCGTGACTTGTCTTTGTGCGACAATCTTTT
 TGTGTTGGTTGTGTGGTGAAGCTACTGGAAGGGACCACACTTGGGCTAGAATCTCGGGTCAT
 AGTTGTGGTCGGTTCCAAGAAGATAAAGAGAAACAAATGGAGAGAGCGAAAAGGGATCTC
 AAGCGGTATATGCATTATCATTAACCGATACAAAGCACATATCGACTCCTCCAAGCTAGAG
 GCTAAGCTTAGTAATAATATTAGTAAAAAGGTGTCTATTTAGAAAAGAGGGAGTTACAA
 CTTAAAGACTTCAGCTGGGCTACCAATGGACTCCATCGGTATTATAGATCAAGACGAGTT
 CTTTCATATTATACCTTTTCGCATTTTACATGTTTGGAGATGAGCTGTTTAAAGATGAG
 ATGAGCTCTGAGGAAAGAGAAATAAAACAAAATCTGTTTGGAGATCAGCAGCAGCAGCTT
 GAGGCTAATGTTGAGAACTTTCTAAGTCTTGGAGGAACCTTTTGATCAATTTGCTGAT
 GATAAGGTCATGCAGATAAGGATTCAAGTCATCAATTTGTCAGTTGCGGTCGATACACTC
 TGCGAAAATATGTATGAATGCATTGAGAATGACTTGTGGGTTCTCTGCAACTTGGCATC
 CACAACATTACTCCATACAGATCAAACGGCATAGAACGAGCATCTGATTTTATAGTTCC
 CAGAATTCCAAGGAAGCTGTTGGTCAGTCTTCGGATTGTGGATGGACGTCAGGCTCGAT
 CAAGCTTTGGAGTCAGGGAAGTCGGAAGACACAAGTTGCTCTTCCGGAAGCGTGCTAGA
 ATAGACGAAAGTTACAGAAACAGCCAAACACCTTACTAGATTTAAACTTGCCAGCGGAA
 GCCATTGAGCGGAAATGAACACTTATCCTTCTTACCTCCCAATAACACCTTTTGTGTC
 AAATAAAGTGTGTTACCCGGATATTATAGCTCTAAACCCAATCCCCTCTGCTTAATTTG
 TCAGTGACCTTACCTAACCTCTTCA
 >G1777 Amino Acid Sequence (domain in AA coordinates:124-247)
 MDDNLSGEEEDYYYSSDQESLNGIDNDESVSIPVSSRSNTVKVITKESLLAAQREDLRRV
 MELLVKEHHARTLLIHYRWDVEKLFVAVLVEKKGKDSLFSGAGVTLLNQSCDSSVSGSSS
 MMSCDICVEDVPGYQLTRMDCGHSFCNNCWTHGFTVKINEGQSKRIICMAHKCNAICDED
 VVRLVSKSQPDLAKEFDRFLLESYIEDNKMVKWCPSTPHCGNAIRVEDDELCEVEECSCG
 LQFCFSCSSQAHSPSCVMWELWRKKCFDESETVNWITVHTKPCPKCHKPVEKNNGCNLV
 TCLCRQSFCLWLCGEATGRDHTWARISGHSCGRFQEDKEKQMERAKRDLKRYMHYHNRYKA
 HIDSCKLEAKLSNNISKVSISEKRELQKDFSWATNGLHRLFRSRRVLSYSYPFAFYMF
 GDELFPKDEMSSEEREIKQNLFFEDQQQQLEANVEKLSKFLEFPDQFADDKVMQIRIQVIN
 LSVAVDTLCENMYECIENDLLGSLQLGIHNITPYRSNGIERASDFYSSQNSKEAVGQSSD
 CGWTSRLDQALESGKSEDTSCSSGKRARIDESYRNSQTTLLDLNLPAAIERK*
 >G1793 (59..1783)
 AGTGATTTATTGATTAAACCCAAACACAAAATAAACAGATTTGACTCAAAAAGAAGAAAAT
 GAATTCTAACAACTGGCTTGGCTTTCTCTTTTACCGAACAACCTTCTTTGCTCCTCA
 TGAATACAACCTTGGCTTGGTCAGCGACCATATGGACAACCTTTTCAAACACAAGAGTG
 GAATATGATCAATCCACACGGTGGAGGAGGAGATGAAGGAGGAGAGGTTCCAAAAGTGGC
 CGATTTTCTCGGTGTGAGCAAAACCGGACGAAAACCAATCCAACCACCTAGTAGCTTACAA
 CGACTCAGACTACTACTTCCATACCAATAGCTTGATGCCTAGCGTCCAATCAAACGATGT
 CGTTGTAGCAGCTTGTGACTCCAATACTCCTAACAACAGTAGCTATCATGAGCTTCAAGA
 GAGTGCTCACAATCTACAGTCACTTACTTTGTCCATGGGGACCACCGCTGGTAATAATGT
 TGTAGACAAAGCTTACAGTCCGAGACCACCGGGGATAACGCTAGCGGTGGAGCACTAGC
 CGTTGTTGAGACGGCCACGCAAGACGTGCATTGGACACTTTCGGACAACGAACCTCGAT
 CTATCGTGGTGTCAAGACATCGATGGACTGGTTCGATATGAGGCTCATCTATGGGATAA
 TAGTTGTAGAAGGAAGGCCAGTCTAGGAAAGGAAGACAAGTTTACTTGGGTGGATATGA
 CAAAGAAGATAAAGCAGCAAGATCATATGATCTAGCTGCACTTAAGTACTGGGGTCTCTC
 AACTACTACTAATTTCCCCATTACAACTACGAGAAAGAAGTAGAGGAAATGAAGCACAT
 GACGAGACAAGAGTTTCGTGGCTGCCATTAGAAGGAAAAGTAGTGGATTTTCGAGAGGCGC
 TTCGATGTATCGAGGAGTTACAAGGCATCACCAACATGGAAGATGGCAAGCAAGGATCGG
 CCGAGTCGCCGGAACAAAGACCTTACTTGGGAACCTTTAGCACTGAGGAAGAAGCAGC
 AGAAGCTTACGATATAGCTGCAATAAAGTTTAGAGGACTTAATGCAGTGACCAACTTCGA
 GATCAACCGGTACGACGTGAAAGCCATTCTAGAGAGTAGCACTCTTCCATCGGAGGAGG
 CGCAGCTAAACGGCTCAAAGAAGCTCAAGCTCTTGAGTCTTCAAGGAAACGCGAGGCGGA
 GATGATAGCCCTTGGTTCAAGTTTCCAGTACGGTGGTGGCTCGAGCACAGGCTCTGGCTC
 CACCTCATCAAGACTTCAGCTTCAACCTTACCCTCTAAGCATTCAACAACCATTAGAGCC
 TTTTCTATCTCTTCAACAATGACATCTCTCATTACAACAACAATGCTCACGATTC
 CTCCTCTTTAATCACCATAGCTATATCCAGACACAACCTTCATCTCCACCAACAGACCAA
 CAATTACTTGCAGCAACAGCTCGAGCCAGAACTCTCAGCAGCTCTACAATGCGTATCTTCA
 TAGCAATCCGGCTCTGCTTCATGGACTTGTCTCTACCTCTATCGTTGACAACAATAATAA
 CAATGGAGGCTCTAGTGGGAGCTACAACACTGCAGCATTCTTGGGAACCACGGTATTGG

TATTGGGTCCAGCTCGACTGTTGGATCGACCGAGGAGTTTCCAACCGTTAAACAGATTA
CGATATGCCTTCCAGTGATGGAACCGGAGGGTATAGTGTTGGACCACTGAGTCTGTTCA
GGGGTCAAACCTGGTGGTGTCTTCTACTATGTGGAATGAGTAAACAAGGATCTCTTCTT
CGGCACAAGGAATGGGT

>G1793 Amino Acid Sequence (conserved domain in AA coordinates: 179-255, 281-349)

MNSNNWLGFPPLSPNNSSLPHEYNLGLVSDHMDNPFQTQEWNMINPHGGGGDEGGEVPKV
ADFLGVSKPDENQSNHLVAYNDSYYFHTNSLMPVQSNVVAACDSNTPNNSSYHELQ
ESAHNLSLTLTSMGTTAGNNVVDKASPSETTGDNASGGALAVVETATPRRALDTFGQRTS
IYRGVTRHRWTGRYEHLWDNSCRREGQSRKGRQVYLGQYDKEDKAARSYDLAALKYWGP
STTTNFPITNYEKEVEEMKHMTRQEFVAAIRKSSGFSRGASMYRGVTRHHQHGRWQARI
GRVAGNKDLYLGTFTSTEEEAAYDIAAIKFRGLNAVTFEINRYDVKAILESSTLPIGG
GAAKRLKEAQALESSRKREAEMLGSSSFQYGGSSSTSGSTSSRLQLQPYPLSIQPLE
PFLSLQNNDISHYNNNNNAHDSSSPNHHSYIQTQLHLHQQTNNYLQQQSSQNSQLYNAYL
HSNPALLHGLVSTSIVDNNNNNGSSSGSYNTAAFLGNHGIGIGSSSTVGSSTEEFPTVKTD
YDMPSSDGTGGYSGWTSSESQSNPGGVFTMWNE*

>G180 (54..629)

GTAATTACGATCTACAACAAGTGACATCGTCGTCGACGACGATTCAAGAGAATATGAACT
TCCTCGTTCCCTTTTGAAGAAACCAATGTCTTAACCTTTTTCTCTTCTTCTTCTCTCTT
CTCTTTCTTCTCTCTTCTTTCCCCATTCACTACTCTTCTCTCCACTACTACTCATGCAC
CTCTAGGGTTTTCTAATAATCTTCAGGGTGGAGGACCTTGGGATCAAAGGTGGTTAATG
ATGATCAGGAGAATTTGGAGGTGGAACATAAATGATGCTCATTCTAATTCTTGGTGGA
GATCAAATAGTGGAAGTGGAGATATGAAGAACAAAGTGAAGATAAGGAGGAACTAAGAG
AGCCAAGATTCTGTTTCCAAACCAAAAGCGATGTTGATGTTCTTGACGATGGCTACAAAT
GGCGTAAATATGGTCAGAAAGTCGTCAAGAACAGCCTTCACCCAGGAGTTATTACAGAT
GCACACACAACAACGTAGGGTGAAAAGAGAGTGGAGCGACTATCGGAAGATTGTAGAA
TGGTGATTACTACTTACGAAGGTCTGTCACAACCACATTCCCTCTGATGACTCCACTTCTC
CTGACCATGATTGTCTCTCTCTTCTTTTAACATCTCTTCTATATATCTATATATAGACAG
TTATATGTGCACATATAGATGTGTGATATATTGCATATTTGATATTGCATGTGTTTTCA
AGAGTATGTCATCAGATGTTATGCATATATTCTTGACTTGTGCTTATAGTATACATATG
TAATAATATATATTGACATTGGTAGTTTCATTCTGTTCAAACAAAAAAAAAAAAA

>G180 Amino Acid Sequence (domain in AA coordinates: 118-174)

MNFLVPFEETNVLTFSSSSSSSLSSPSFPIHNSSTTTTHAPLGSNNLQGGGPLGSKV
VNDQENFGGGTNNDAHSNSWRSNSGSDMKNVKIRKLRPRFCFQTKSDVDVLDG
YKWRKYGQKVKNLSLHPRSYYRCHNNCRVKRVERLSEDCRMVITTYEGRHNHIPSDDS
TSPDHDCLSSF*

>G192 (63..959)

CTTTTTTCTCTTCTCTCCTCAGAGATTGGAAGCTTTTTGTCTCCCTGAGTAACCAATT
CAATGGCCGACGATTGGGATCTCCACGCCGTAGTCAGAGGCTGCTCAGCCGTAAGCTCAT
CAGCTACTACCACCGTATATTCCCCCGGCGTTTCATCTCACACAAACCTATATTACCG
TCGGACGACAAAGTAATGCCGTCTCCTTCGGAGAGATTGAGATCTCTACACACCGTTCA
CACAAGAATCTGTCGTCTCTTCGTTTTCTTGATATAAACTACCCAGAAGAACCTAGAAAGC
CACAGAACCAGAAACGTCCTCTTCTCTCTCTGCTTCTTCCGGTAGCGTCACTAGCAAAC
CCAGTGGCTCCAATACCTCTAGATCTAAAAGAAGAAAGATACAGCATAAGAAAGTGTGCC
ATGTAGCAGCAGAAGCTTTAAACTCCGATGTCTGGGCATGGCGAAAGTACGGACAGAAAC
CCATCAAAGGTTACCATATCCAGAGGATACTACAGATGTAGTACATCAAAGGTTGTT
TAGCCCGTAAACAAGTGGAGCGAAATAGATCCGACCCGAAGATGTTTATCGTCACTTACA
CGGCGGAGCATAATCATCCAGCTCCGACACACCGTAATTCTCTCGCCGGAAGCACACGTC
AGAAACCATCCGATCAACAGACGAGTAAATCTCCGACGACCACTATTGCTACTTATTCAT
CGTCTCCGGTGACTTCAGCCGACGAATTTGTTTTGCCTGTTGAGGATCATCTAGCGGTGG
GAGATCTTGACCGGAGAAGAAGATCTGTTATCTTTGTGCGGATACGGTGGTTAGCGATGATT
TCTTCGATGGGTTAGAGGAATTCGAGCCGGAGATAGCTTTCCGGGAACTCGGCTCCGG
CGAGTTTTGATCTCTCTTGGGTTGTGAACAGTGCAGCCCACTACCAACGGAGGAATATGAT
TAGATTACGACGGCTTAGAATACTCTTATTAGGACAGATTTATAGGATTAAGGAATTATT
CTCGGAGCATATGTAAAAATAGGATAAAAGAAAATGTTCTTTGTTACTTTTTTTTCGGGT
TTCTTCTATTGTTTCTAAACATCTTAGAAAAAATTTAATTGTATATTCCTTAAGCTCGA
TACATCTTGTTTTTAAAAAAAAAAAAAAAAA

>G192 Amino Acid Sequence (domain in AA coordinates: 128-185)

AAAAGGTCTTCTTGGCCATGGATACTTGTGCTCTAGTAATCCATCAGTCTCTGTCTCGCA
 TCAAACCTTTCTCCTCCCAAATCTTCTTCTTCTTCTTCTGCTTTCTCCCTGAATCCT
 TACCGATCAGACGGATCGAGCTGTGTTTCCGAGGAGCTATATGTGCCGCCGTACAAAGAA
 ACTACGAAGAAACGACCTCCTCCGTGGAAGAGGCAGAGGAAGATGATGAGTCATCATCAT
 CGTACGGGAAGTGAAACAAGATCATTTGGAAGCCGAACGGCGGGGGAAGGAGCCATGGAGT
 ACCTTATCGAGTGGGAAGGACGGCCATTCTCCGTCGTGGGTTCATCGAGCTACATCGCAG
 CAGACGTAGTGTGCGAGTACGAGACACCTTGGTGGACGCGAGCTAGAAAGCCGACGAGC
 AGGCCCTGTACAGCTCCTGGAGGACCGAGACGTGATGCCGTGGACGAAACGGCCGGA
 CGGCTCTGCTTTTCTGGCAGGTCTGGGGTCGGACAAGTGCCTAAGGCTTCTGGCGGAGG
 CTGGAGCCGATCTCGACCACCGAGACATGAGGGGAGGCTTGACGGCGCTGCACATGGCGG
 CTGGTTACGTGAGGCCGGAGGTGGTGGAGGCGCTGGTGGAGCTGGGAGCTGATATTGAAG
 TGAAGACGAGAGAGGGGTTAACGGCGTTGGAATAGCGAGGGGAGATTCTGAAGACGACGC
 CGAAGGGGAATCCGATGCAGTCTCGGAGGAGAATTGGTTAGAGAAAGTGATCAATGTCC
 TGAAGGACAAGTGTTCCGAGTACGCCAGGTGGATGAGATCGTAGAGAAACGGGGAAAG
 GCAAAGACGTTGAATATCTGGTCAGATGGAAGGACGGTGGAGATTGCGAGTGGGGTGAAG
 GTGTACACGTGGCGGAAGATGTGGCTAAGGACTACGAGGATGGGCTGGAGTACGCTGTAG
 CGGAGAGTGTGATCGGGAAGAGGGTGGGAGACGATGGAAGACCATCGAGTATCTTGTCA
 AATGGACTGATATGTCTGATGCCACTTGGGAGCCTCAGGACAATGTGACTCTTACTCTTG
 TTCTACTCTACCAACAACAACCAATGAATGAATGATTGATTTTGATGATTACATTCT
 TCTCAATTGTCTTCTTCTCATATGTGTGGTTCTATGACCGGTTCCGTTGGTACGTAC
 CGGTACATTTTCTATTTCTTTTAAAGATGTGATCTTGATGGTTTTTGGCCATTTTGGGACA
 CTATTTGATTTTATATCCATGCTTTGAATTTTGCCTCCCTTTTTGGGGAGATTATGAAG

MDTCALVIHQSLSRIKLSPPKSSSSSSSAFSPESLPRIIRIELCFRGAICAAVQRNYEETT
SSVEEAEDDESSSYGEVNKIIGSRTAGEGAMEYLI EWKDGHS PSWVPS SYIAADVSE
YETPWWTAAARKADEQALSQ LLED RDVDAVDENGRTALLFVAGLSDKCVRL LAEAGADLD
HRDMRGGLTALHMAAGYVRPEVVEALVELGADIEVEDERGLTALELAREILKTPKGNPM
QFGRRIEGLKVINVLEGQVFEYAEVDEIVEKRGKGKDVEYLVRWKDGGDCEWVKGVHVAE
DVAKDYGLEGYAVAESVIGKRVGDDGKTI EYLVKWTMSDATWEPQDNVDSTLVLLYQQ
QPMNE*

ATGAGAAAAGTATGTGAGCTTGATATAGAGCTAAGTGAAGAGGAAAGAGACCTACTAACA
ACTGGATACAAGAATGTCATGGAGGCTAAGAGAGTTTCATTGAGAGTAATATCATCCATT
GAAAAAATGGAAGACTCGAAAGGAAACGACCAAAATGTGAAACTGATAAAAGGACAACAA
GAAATGGTTAAATATGAGTTTTTCAATGTTTGTAATGACATTTTGTCTCTCATTGATTCT
CATCTCATACCATCAACTACTACTAATGTGCAATCAATTGTCCTTTTTTAACAGAGTGAAA
GGAGATTATTTTCGATATATGGCAGAGTTTGGTTCTGATGCTGAACGTAAGAAAATGCA
GATAATTCTCTAGATGCATATAAGGTTGCAATGGAATGGCAGAGAATAGTTTAGCACCC
ACCAATTGGTTAGACTGGAGTTGGCTTTAAATTTCTCGATATTCAATTATGAGATCCAT
AAATCTATTGAAAGCGCATGTAAATTGGTTTAAGAAAGCTTACGATGAAGCAATCACTGAA
CTCGATGGCCTTGACAAGAATATATGCGAAGAGAGCATGTATATCATAGAGATGCTTAAA
TACAATCTTTCTACGTGGACTTCAGGCGATGGTAATGGTAATAAGACAGCGGTTAG

GGAAACCCTAATTTCCGCAAATTCCTATGAAGCGTATTATCAGAATCTCATTACCGGACG
CAGAAGCCACCGATTCTTCTAGCGACGAAGACACGGAGGAGCGTGGAGGAGCATCCAGA
CTCGGCGCGGTGGGAAACGCCCTCGTTAAAGAGATCGTAATCGATCCTTCCGATTCCGCCG

219

ALSSSFPSF*

>G265 (280..1317)

CTTTGGTCTTGAAGCCAAATCAAACCTTTCTTCAATCCTCAAATTTTCGAAAATTTTC
TCTTTTGCTTTACGTTCTCTCAATTCCTATTGTAAGAAAGTTTGTTCCTTTAATCAATC
AAATCAAAGAGACTTTTGAAGATTGTTTCCCAATTTGCGTCAATCGGGATCGAGTCAAAT
CTGAAATCTTCTCCACTCATCTGACTATAAGACTTAATCAAGGGACTTTTGTTCGG
GTTTGGTTTTAAACGCTCTTGGATTGGAAGTGGTTAAGGTATGGATGAAAATAATGGAGGT
TCAAGCTCACTTCCACCTTTCTTACTAAAACATATGAAATGGTTGATGATTCTTCTTCT
GACTCGGTGCTTGGCTTGGAGCGAAAACAACAAAGCTTCATCGTCAAGAATCCAGCAGAG
TTTTCAAGAGACCTTCTTCCGAGATTCTTCAAGCATAAGAATTTCTCAAGTTTCATCCGT
CAGCTTAATACATATGGTTTTTCGAAAAGTAGATCCTGAGAAATGGGAATTTCTGAATGAT
GATTTTGTAGAGGTGACCTTACCTTATGAAGAACAATTCATAGACGAAAACCGGTTTCAT
AGCCACTCGTTAGTGAATCTACAAGCGCAAAATCCTTTGACGGAATCAGAAAGACGGAGC
ATGGAGGATCAGATAGAAAGACTGAAAAATGAGAAAGAAGGCCTTCTTGGCGAGTTACAG
AACCAAGAGCAAGAACGGAAAGAGTTTGAGCTGCAAGTAACGACATTGAAAGATCGGTTA
CAACATATGGAACAACATCAGAAATCAATAGTGGCATATGTTTCACAGGTTTGGGAAAA
CCAGGACTTTCACTAAACCTCGAAAACCATGAGAGAAGAAAAAGAAGATTCAAGAGAAC
TCTCTTCTCCAAGCAGTTACACATAGAACAGGTGCAAAAGTTAGAATCTTCGCTAACG
TTTTGGGAGAATCTTGTATCGGAATCATGCGAGAAGAGCGGTTTGCAGTCATCAAGCATG
GATCATGATGCAGCTGAGTCAAGTCTAAGTATTGGCGATACACGACCCAAATCATCGAAG
ATTGATATGAACCTCAGAGCCGCCCGTTACCGTTACTGCGCCTGCTCCAAAAACAGGCGTT
AACGATGACTTTTGGGAACAATGTTTGACAGAGAACCCTGGATCAACCAGCAACAAGAA
GTTCAAGTCAGAGAGAAGAGATGTCGGTAATGATAATAATGGTAATAAGATTGGAAATCAA
AGGACGTATTGGTGGAAATTCAGGGAATGTAAATAACATTACAGAGAAAGCTTCTTGACAT
GAATGAGGTTTTTGTAAATAAGTTTTCTTTTGGTTCCACTGAGATTATTGTATGTGTTCA
TTATTTATTACTCTGTTTCTGTAAAAACAAATCTCTCTATTGTTTGAGGCAGGAGTGACA
TAAATGCATATGCAGAATTGGTTTCAAAA

>G265 Amino Acid Sequence (domain in AA coordinates: 11-105)

MDENNGSSSLPFLTKTYEMVDDSSSDSVVAWSENNKSFIVKNPAEFSRDLLPRFFKHK
NFSSFIRQLNTYGRKVDPEKWEFLNDDFVRGRPYLMKNIHRRKPVHSHSLVNLQAQNPL
TESERRSMEDQIERLKNKEGLLAELOQEQERKEFELQVTTLDKRLQHMEQHOKSIVAY
VSQVLGKPGLSLNLNLEHRRKRRFQENSLPPSSSHIEQVEKLESSLTFWENLVSECEKS
GLQSSMDHDAESSLSIGDTRPKSSKIDMNSEPPVTVTAPAPKTGVNDDFWEQCLTENP
GSTEQQEVQSERRDVGNDNNGNKIGNQRTYWWNSGNVNNITEKAS*

>G2792 (1..960)

ATGGATCATCATCATCATAGCATCAAGAAATTCATCAACAACATCAGAATTACCATCA
TTCGAGCCAGCGTGCCATAACGGTAATGGTAACGGTTGGATCTATGACCCAAATCAAGTT
AGGTACGATCAAAGTAGTGACCAACGGCTGTCAAAGTTGACGGATCTTGTAGGCAAGCAC
TGGTCAATTGCACCACGAATAATCCCGACATGAACCATAACCTTCATCATCACTTCGAT
CATGATCATTCTCAAACGACGACATTTCTATGTACAGACAAGCCTTGGAGGTGAAAAAT
GAGGAAGATCTTTGTTACAATAATAGGCTCAAGTGGTGGTGGTTCTTGTTCATGATCCT
ATAGAAAGTTCTAGAAGTTTCTTGTATATAAGGTTAAGTAGGCCATTAACGATATTAAT
CCGTCAATTAAGCCATGCTTTAAGGCCTTAAACGTATCCGAGTTTAAACAAGAAAGACAT
CAAACGGCATCTCTGGCAGCAGTGAGACTGGGAACAACAACCGTGGAAAAAGAAGAGA
TGTGAAGAAATTTCCGATGAGGTTTCAAAGAAGGCAAGTGCAGTGAGGGCTCTACACTT
TCGCCAGAGAAGGAATACCCAAAGCCAACTTCGAGACAAGATCACGACTCTACAGCAA
ATTGTGTCTCCCTTTGGAAAGACTGATACTGCTTCTGTGCTTCAAGAGGCCATCACTTAC
ATAAATTTTTATCAAGAGCAAGTTAAGCTGCTAAGCACTCCTTATATGAAGAATTCATCA
ATGAAGGATCCATGGGGGGGATGGGACAGAGAAGATCACAACAAAAGGGGACCGAAGCAT
CTTGATCTAAGGAGTAGAGGGCTTTGTTTGGTTCTTATTTCATATACCCCAATCGCATAC
CGCGATAACAGTGCAACTGACTACTGGAATCCACGTATAGAGGTTCTTTGTATCGTTAG

>G2792 Amino Acid Sequence (domain in AA coordinates: 190-258)

MDHHHHIASRNSSTSELPSFEPACHNGNGNGWIYDPNQVRDQSSDQRLSKLTDLVGKH
WSIAPPNNPDMNHNLHHHFDHDSQNDISMYRQALEVKNEEDLCYNNNGSSGGGSLFHDP
IESSRSFLDIRLSRPLTDINPSFKPCFKALNVSEFNKKEHQTASLAAVRLGTTNAGKKR
CEEISDEVSKKAKCSEGSTLSPEKELPKAKLRDKITTLQQIVSPFGKTDASVLQEAITY
INFYQEQVKLLSTPYMKNSSMKDPWGGWDREDHNRGPKHLDLRSRGLCLVPISYTPIAY

RDNSATDYWNPTYRGSLYR*

>G2830 (1..903)

ATGTCTTCCATCCCAAATAGGTTCAATATTTATGGTGGTGATACCACAAACCATCGTGAA
TCGCTTCCCATCGAAATGAATCACAACCTCTCGAATGGTTCGATCCATGTTTCATTACATCT
GATCGCATGAATCATAGAGATTTGTTTTCTTCTCCTCCTTCTTCTTCTTATCAAAAT
TCACATATCTCTTCATCTTCTGTTGGGTTTAATAATTACATATGACTTATCATATGCTG
AAAAGAAATTATGATTCTGTTTCCCGTGCTGATTATTTCTCTACTAAAGATCATTCTCAT
TTTACTCAAGTATCTTTCACCTCAAACCATCACAAATAAGTATACTACTATTGTTCCCTTCC
AATATATTTGACACTGTTCACTATGATATTGGTTCGTGTCAAACGTGCCATAGATTTTAGA
AATATTTGGAATCTTAAATCTCATCTTCCAAAAAATTTAATAGGCAATGCGAGATTTTG
AATCCTACCCCTCTTAATATCGTCTTTCCGCACCAGGATTCAGCTGATCGTCAACATTTA
GACATTATTTTCTCGTCATCAAAGCACAAACCATGTTTTCGAAGATGGTTCGATCCTTGAAG
AAAATTTCCGAACCAACCAATCTGTTTGAAAAATCTAATCTTATGATTCTCAAGAAGAT
GAGAAATCGATGCTTATCAATATGATGGTTCGTACACATAGTCTACCGTATACGAAATAC
GGTCCATATACATGTCCAGGTGTAACGGTGTGTTTGATACTTCTCAAAAAATTTGCTGCA
CATATGTTATCTCACAATAATGAGACGGACAAAGAAAGAGACCAAAGATTTTCGTGCA
AGAAATAAAAAACGATATCGTAAGTTTATGGACAGTCTTAAATATCAAAACAGAAGATA
TGA

>G2830 Amino Acid Sequence (domain in AA coordinates:245-266)

MSSIPNRFNIYGGDTTNHRESLPIEMNHNHRDLFSSPPSFSSYQN
SHISSSSVGFNNSHMTYHMLKRYDSVSRADYFSTKDSHFTQVSFTQITINKYTTIVPS
NIFDTVHYDIGRVKRAIDFRNIWNPKSHLPKKFNRQCEILNPTPLNIVFPHQDSADROHL
DIIFFSSSKHNHVFQDGRSLKKISEPTNLFKESNSYDSQEDEKIDAYQYDGRTHSLPYTKY
GPYTCPRCNGVFDTSQKFAAHMLSHYNNETDKERDQRFARNKKRYRKFMDSLKISKQKI

*

>G286 (94..2454)

TGCAATTTCTCTCGACCAAACCTAATTTTCAGGTTTGGGGTTTTCCTTCTTTCCTGTC
AATTTTGATGAACTTGTGATTAGTATGAAATGAATGCTAATGAGCAAACCTCGATCC
GCCAATGGCATTGGCAATGGCAATGGTGAAGTCTATTCCCGGGATTCCAGATGACTTACGG
TGCAAGAGATCGGATGGTAAACAGTGGAGATGCACTGCAATGTCCATGGCTGATAAGACT
GTTTGTGAGAAGCACTACATCCAAGCAAAGAAGCGGGCGGCTAATCTGCTTTCAGGGCG
AACCAGAAGAAAGCGAAAAGCGGATCATCGTTAGGCGAAACAGATACGTATTTCGGAAGGG
AAGATGGATGATTTTCAGTGTACAGTACCAGCATTGACCACTATAATAACGGTCTTGCC
TCTGCTTCCAAGAGTAATGGTAGACTAGAGAAGAGACATAATAAAGCCTGATGCGGTAC
TCGCCCGAGACACCGATGATGAGGAGTTTCTCTCCACGTGTTGCAGTGGATTTGAATGAT
GACTTGGGTAGAGATGTTGTAATGTTTGAAGAGGGCTACAGATCTTATAGACACCACCA
TCTGTTGCTGTTATGGATCCGACACGAAACAGATCACACCAAAGCACCAGTCTATGGAA
TACTCAGCAGCAAGCACAGATGTGTCTGCAGAGTCTTTGGGGGAAATCTGCCATCAATGC
CAGAGAAAAGATAGAGAGAGAATCATTTCTTGCCCTCAAATGCAATCAAAGAGCCTTCTGC
CACAAATGTCTATCGGCAAGGTACTCGGAGATATCACTTGAAGAAGTCGAGAAAGTTTGC
CCTGCATGTCGTGGCTTGTGTGATTGCAAATCTTGCCCTGCGTTTCAGATAATAACAATAAG
GTTCCGATCCGGGAAATACCGTTTTTGGACAAGTTGCAGTATCTTTATCGTCTATTATCA
GCTGTCCTACCAGTCATAAAGCAGATCCATCTTGAACAATGTATGGAAGTTGAACTAGAG
AAGAGGCTTCTTGAAGTTGAGATTGATCTTGTTCAGGGCAAGATTGAAAGCAGATGAGCAG
ATGTGCTGCAACGTGTGTCCGATACAGTGTGTTGACTACTACCGTCACTGTCCGAAGTGC
TCATATGACCTTTGCCTGAGATGCTGTCAAGATCTACGGGAAGAGTCTTCAGTGACGATT
AGTGGGACTAACCAAAACGTACAAGATAGAAAAGGAGCTCCCAAACCTAAACTAACTTT
TCATACAAGTTTCTGAGTGGGAAGCCAACGGTGTATGGGAGCATCCCTTGCCCTCCTAAG
GAGTATGGAGGCTGCGGTTACATCTTTGAATCTTGCCCGCATTTTCAAGATGAATTGG
GTTGCAAAGCTTGTGAAAAATGCTGAGGAGATTGTTAGTGGCTGCAAATTATCTGATCTT
CTGAACCTGATATGTGTGATTCAAGATTCTGCAAATTTGCTGAGAGAGAAGAGAGCGGT
GACAACCTACGTGTACAGCCCGTTCGCTTGAAACGATTAATACTGATGGAGTAGCTAAGTTT
GAGCAACAATGGGCGAGAGGGTTCGCTTGTACTGTGAAATGGTACTTGATGACTCATCT
TGCTCTAGATGGGATCCTGAGACTATTGAGAGGATATAGACGAGCTTTTCGGACGAGAAA
CTGAGAGAACATGATCCATTCTTGAAGGCCATTAATTGCTTGGATGGTTTAGAGGTTGAT
GTAAGACTTGGGGAGTTTACAAGAGCATATAAAGATGGAAGAACCAAGAGACAGGTCTT
CCGCTATTGTGGAAGTTAAAGGACTGGCCGAGCCCAAGTGTCTCCGAGGAGTTTCAATTTT

TACCAAAGACCTGAGTTTATCAGAAGTTTCCGTTTCTCGAGTACATTCATCCCCGGTTA
GGCCTTCTGAATGTTGCAGCCAAGTTACCTCATTACTCGCTCCAAAACGATTTCAGGTCCA
AAGATTTATGTGTCTTGTGGGACGTACCAAGAAATCAGTGTGGCGATTTCATTGACTGGT
ATTCACTACAACATGCGTGACATGGTATACCTATTGGTGCACACGTCTGAAGAAACAACA
TTCGAAAGGGTGAGAAAAACAAACCTGTTCCAGAGGAACCTGACCAGAAGATGAGCGAA
AATGAGTCACTTCTTAGCCCTGAGCAGAAATTAAGGGACGGAGAGTTACATGATCTATCA
CTTGGTGAAGCCAGTATGGAGAAGAATGAACCTGAGTTGGCGTTGACTGTGAATCCAGAG
AACTTAACGGAACCGGTGACAACATGGAATCTTCTTGACATCTTCATGTGCAGGAGGA
GCCCAGTGGGATGTCTTTCGACGCCAAGACGTCCCAAAGTTGTCCGGGTATTTGCAGAGA
ACATTCCAGAAGCCTGATAATATCCAGACTGATTTTGTAAAGCCGTACCTGCTAATTCAA
TAAATGAAGTGTGTAAAGTCTGTATGTGGAATGATTGAGTTTCCTAGTTTGTCTTACTCT
GGTTTCAGGTGTACGCCCCGTGTATGAAGGATTGTCTTTAAATGAACACCACAAGAGAC
AACTAAGAGACGAGTTTGGAGTTGAGCCATGGACATTTGAGCAACATCGTGGTGAGGCTA
TCTTCATTCCGGCTGGATGTCCGTTCCAAATCACTAATCTTCAGTCGAATATTCAGGTGG
CACTTGACTTCTTGTGCCCTGAAAGCGTTGGAGAGTCAGCAAGACTAGCTGAAGAAATCC
GGTGTTTACCAAACGACCACGAGGCAAACTTCAGATTCTAGAGATTGGAAGATATCAT
TATACGCAGCTAGCTCAGCCATTAAAGAGGTTGAGAACTGGTCTTGGATCCAAAGTTTG
GAGCAGAGCTTGGATTTGAAGACTCTAACTTAACCAAAGCAGTCTCTCACAACCTTAGACG
AGGCAACCAAGCGGCC

>G286 Amino Acid Sequence (domain in AA coordinates: TBD)
MNAEQTRSANGIGNNGESIPGIPDDLRCRSDGKQWRCTAMSMADKTVCCKHYIQAKK
RAANSAFRANQKAKRRSSLGETDTYSEGKMDDFELPVTSIDHYNNGLASASKSNRLEK
RHNKSLMRYSPETPMRFSFPRVAVDLNDLGRDVMFEEGYRSYRPPSVAVMDPTRNR
SHQSTSPMEYSAASTDVSABSLGEICHQCQRKDRERIISCLKCNQRAFCHNCLSARYSEI
SLEEVEKVCAPACRGLCDCKSLRSDNTIKVRIREIPVLDKLYLYRLLSAVLPVIKQIHL
EQCMEVELEKRLLEVEIDLVRARLKADEQMCCNVCRIPVVDYRHCPCNSYDLCLRCCQD
LREESSVTISGTNQNVDKRGAPKLKLNFSYKPEWEANGDGSIPCPKEYGGCGSHSLN
LARIFKMNWVAKLVKNAEEIVSGCKLSDLLNPDMDCSRFCFAEREESGDNYVYSPSLET
IKTDGVAKFEQQWAEGRILVTVMVLDDSSCSRWDPETIWRDIDELSDKLREHDPFLKAI
NCLDGLLEVDRLEGEFTRAYKDGKNQETGLPLLWLKLDWSPSPASEEFIFYQRPEFIRSF
FLEYIHPRLLGLLNVAAKLPHYSLQNDSGPKIYVSCGTYQEISAGDSLTIHYNMRDMVYL
LVHTSEETTFERVRKTKPVPEEPDQKMSSENLSPQKLRDGLHDLISLGEASMEKNEP
ELALTVNPNELTENGDNMESSCTSSCAGGAQWDVFRQDVPKLSGYLQRTFOKPDNIQTD
FVSRTC*

>G291 (124..1197)
CAAGAACCCAAAGATCTCTCTCTATTTGTTTGCCTTCTTCTTTCTTTCTGACTCAAACCC
TCAAATCAATTCTCGCGATTAAGCAAAACCCTAGATTTATTCTACTCTTCGAAGTCGATT
TCAATGGAAGGTTCTCTCGTCAGCCATCGCGAGGAAGACATGGGAGCTAGAGAACAACATT
CTCCAGTGGAACCAACCGATTACGCTCCGACAGTATATTCCACTACGACGACGCTTCA
CAAGCCAAAATCCAGCAGGAGAAGCCATGGGCCTCCGATCCTAACTACTTCAAGCGCGTT
CACATCTCAGCCCTTGCTCTTCTCAAGATGGTGGTTACGCTCGCTCCGGTGGCACAATC
GAGATCATGGGTCTTATGCAGGGTAAACCGAGGGTGATACAATCATCGTTATGGATGCT
TTTGCTTTGCCTGTTGAAGGTACTGAGACTAGGGTTAATGCTCAGTCTGATGCCTATGAG
TATATGGTTGAATACTCTCAGACCAGCAAGCTGGCTGGGAGGTTGGAGAACGTTGTTGGA
TGGTATCACTCTCACCTTGGGTATGGATGTTGGCTCTCGGGTATTGATGTTTCGACACAG
ATGCTTAACCAACAGTATCAGGAGCCATTCTTAGCTGTTGTTATTGATCCAACAAGGACT
GTTTCGGCTGGTAAGGTTGAGATTGGGGCATTGAGAACATATCCAGAGGGACATAAGATC
TCGGATGATCATGTTCTGAGTATCAGACTATCCCTCTTAACAAGATTGAGGACTTTGGT
GTACATTGCAACAGTACTACTCATTGGACATCACTTATTTCAAGTCATCTCTCGATAGT
CACCTTCTGGATCTCTTGTGAACAAGTACTGGGTGAACACTCTTCTTCTTCCCACTG
TTGGGCAATGGAGACTATGTTGCCGGGCAAATATCAGACTTGGCTGAGAAGCTCGAGCAA
GCGGAGAGTCAGCTCGCTAACTCCCGGTATGGAGGAATTGCGCCAGCCGGTCACCAAAGG
AGGAAAGAGGATGAGCCTCAACTCGCGAAGATAACTCGGGATAGTGCAAAGATAACTGTC
GAGCAGGTCCATGACTAATGTACAGGTTATCAAAGACATCTTGTTCATTCCGCTCGT
CAGTCCAAGAAGTCTGCTGACGACTCATCAGATCCAGAGCCCATGATTACATCGTGAAGT
TGGTCTATTCTTTTGTGTTTGGCTGCGGAAATTGACTATCGGTTTGACCCGGTTTATGA
GGCAATGCCATTGTTCCCTATATCTCTAGTGTAGTATCTGCTTCAGACAAAGATCTTTG

GGTTATTAAATGACATTAACATAAAAAAA
>G291 Amino Acid Sequence (domain in AA coordinates: 132-160)
MEGSSSAIARKTWELENNILPVEPTDSASDSIFHYDDASQAKIQQEKPWASDPNYFKRVH
ISALALLKMOVHARSGGTIEIMGLMQKTEGDTIIVMDAFALPVEGTETRVNAQSDAYEY
MVEYSQTSKLAGRLENVVGYHSHPGYGCWLSGIDVSTQMLNQQYQEPFLAVVIDPRTV
SAGKVEIGAFRTYPEGHKISDDHVSEYQTIPLNKIEDFGVHCKQYSLDITYFKSSLDLSDH
LLDLLWNKYVWNTLSSSPLLNGNDYVAGQISDLAEKLEQAESQLANSRYGGIAPAGHQRR
KEDEPQLAKITRDSAKITVEQVHGLMSQVIKIDILFNSARQSKSADDSDDPEPMITS*
>G427 (49..1230)

TTTCCCTCTCCGAAACAGAAATTCAAAAACAAATTCAACACGAAAACGATGGCGTTTCAT
AACAAATCACTTTAATCATTTACCGACCAACAACAACATCAGCCTCCTCCTCCGCCGCAA
CAGCAGCAGCAACAACATTTTCAAGAATCAGCACCCCTAATTGGCTCCTCCGCTCCGAC
AACAACTTCTCAATCTCCACACAGCTGCCACAGCCGCCGCTACAAGCTCCGATTCTCCT
TCTTCCGCCGCCGCTAACCACTGGCTCTCACGATCCTCATCCTTCTTCCAACGAGGCAAC
ACCGCAAACAACAACAACGAAACATCCGGTGACGTCATCGAAGACGTTCCCGCGCGGA
GAGGAGTCAATGATCGGAGAGAAGAAGGAGGCGGAGAGGTGGCAGAATGCGAGACACAAG
GCGGAGATACTGTCTCATCCACTATACGAGCAACTTTTGTGCGGCACACGTGGCGTGCCTG
AGGATCGCAACGCCGGTGGATCAGCTTCCGAGGATAGACGCACAGCTTGCTCAGTCTCAA
AACGTCGTGGCTAAGTACTCAACTTTAGAAGCCGCTCAAGGACTCCTCGCCGGCGATGAC
AAGGAGCTTGACCACTTCATGACGCATTATGTACTATTGCTTTGCTCTTTCAAAGAACAA
CTGCAACAGCATGTTTCGTGTTTCATGCAATGGAAGCTGTTATGGCCTGTTGGGAGATTGAA
CAGTCGCTTCAAAGTTTACAGGAGTATCTCCTGGTGAAGGCACAGGAGCAACAATGTCT
GAGGATGAAGATGAGCAAGTAGAGAGTGATGCTCATTTGTTTGATGGAAGCTTAGATGGG
TTAGGGTTTGGTCTCTAGTTCCCACTGAGAGCGAGAGATCTTTGATGGAACGAGTCAGA
CAAGAACTCAAACATGAACCTCAAGCAGGTTTACAAGGAGAAAATTGTGGACATAAGAGAG
GAGATACTGAGGAAGAGAAGAGCTGGAAAATTACCAGGAGACACCACCTCTGTTCTCAA
TCATGGTGGCAATCTCATTTCTAAGTGGCCTTACCCTACTGAGGAAGATAAGGCGAGGTTG
GTGCAGGAGACGGGTTTGCAGCTCAAACAGATAAACAATTGGTTCATCAATCAAAGAAAG
AGGAATTGGCATAGCAATCCATCTTCTTCTACCGTCTCAAAGAATAAACGCCGAAGCAAT
GCAGGTGAAAACAGCGGAAGAGACCGTTGAGATCAAGCTTGATGTAGAGATCCAAAAGC
TTTATAGAAAGGTGGAGGCATGAAGACAAAGAATTCTTACACAACAAACGTAGGACGTAA
TTTTGTGCCAGTACATGGTATGGCTTTTCATATTTGGTAATGATTAGGGCCACACAAAATT
AAACCCCAAAGCATGATTTGTAATATGAGGTTTTAGATGGACTTTATGATAGGATCGTCA
GTCTTCACTGCCATCTCCATTCTCCACCATCAATCCATCATTATATCTTTGTGAAAAAAA
A

>G427 Amino Acid Sequence (domain in AA coordinates: 307-370)
MAFHNNHFNHFTDQQQHQP PPPPQQQQQHQHFQESAPPNWLLRSDNNFLNLHTAATAAATS
SDSPSSAAANQWLSRSSSFLQRGNTANNNNNNETSGDVIEDVPGGEESMIGEKKEAERWQN
ARHKAEILSHPLYEQLLSAHVACLRIATPVDQLPRIDAQLAQSONVVAKYSTLEAAQGLL
AGDDKELDHFMTHYVLLLCFKEQLQQHVRVHAMEAVMACWEIEQSLQSFQVSPGEGTG
ATMSEDEDEQVESDAHLFDGSLDGLGFGPLVPTESERSLMERVQRQELKHELKQGYKEKIV
DIREEILRKRRAGKLPDPTTSVLKSWWQSHSKWPYPTEEDKARLVQETGLQLKQINNWF
NQRKRNWHSNPSSSTVSKNRRSNAGENSGRDR*

>G509 (122..1054)
CTTCCCTCTTTGCTAATAAACTTTTCTTTGAACCTTACACGCCTTGTTGATATTACTCTC
TTAAATATATATTTTCGTACATTAACACAGACATATATAAAGCTAAAGATTCTTCACGT
AATGGGTTTGAAAGATATTGGGTCCAAATTGCCACCGGGGTTTCGATTTTCATCCAAGTGA
TGAAGAGTTGGTTTGTCTATTATCTTTGCAACAAGATTAGGGCCAAATCTGATCATGGTGA
TGTTGATGATGATGATGATGATGTTGATGAAGCTTTGAAGGGTCTACTGATCTTGTGGA
GATTGACTTGATATCTGTGAGCCATGGGAGCTTCTGATGTGGCAAAGTTAAACGCAAA
GGAATGGTACTTCTTTCAGTTTCCGTGATCGAAAGTATGCTACTGGATATCGCACGAACAG
AGCGACAGTAAGCGGATACTGGAAAGCAACAGGAAAAGATCGAACGGTGATGGATCCACG
TACAAGGCAATTGGTAGGGATGAGAAAAACACTAGTGTCTACAGAAACAGAGACCAAA
TGGGATCAAACTACTTGGATCATGCACGAGTTCCGTCTTGAGTGTCTAACATCCACA
TAAGGAAGACTGGGTCTTGTGACAGTGTTCACAAAGGCAGAGACTCATCGTACAAGA
CAATAATTATTATAACAATGATAATCAGACGCAAGGCTTGAAGTTAATGACGCTCCGGA
TCTTAATTACAACAATCAGTTGCCACCTTTGCTATCATCCCTCCTCATAATCATCAACA

TGAGAAGATGAAAATCCAAGTTTGTGATCAGTGGGAGCAGCTAATGAAGCAGCCTTCAAG
GACCACCGGCCACCCCTATCATCACCATTGTTCATCATCAAACCATAGCATGTGGTTGGGA
GCAGATGATGATCGGTTGCTGTTCATCACCTTCGAGTCATGGCCCTGATCACGAGTCCTT
TGCTAAATTTGCTTTACCGTCGACAATAACAACAGTGTCAACATCAGTGGTGATCATCAT
CAGAATTATGAGAAGATTTTGTGTGTCATCACTAGACATGACGAGTTTGGATCACGACAAG
ACATGTATGGGATCATCATCGGATGGTGGTATGGTCTCTGATCTTCACATGGAATGTGGT
GGATTGAGTTTGGAGACCGAAAATATCCTCGCTTTCCAATGAACATAATTCAAGGGGTTT
GCCAATTTGTTGATTTCGTGAATTATACAAACATTTTATCTATAGATTTATCACATTATCA
AACATGTAAGTTGTGTGGCATTGGGTATAGGGTTTGTGTGATTCTAGGTTTATAGGACG
ATGTATGTTGTTATATTTAGCGTGTTTTAGGATTTATTCTCATTTTAAATTTATATGAA
AACCATTACTATGAATACAATTAGTTTCTTTGTGTAAATAATATTTTAGATTATCAA
AAAAAAAAAAAAAA

>G509 Amino Acid Sequence (domain in AA coordinates: 13-169)
MGLKDIGSKLPPGFRFHPSEELVCHYLCNKIRAKSDHGDVDDDDDDVDEALKGSTDLVE
IDLHICEPWELPDVAKLNAKEWYFFSFRDRKYATGYRTNRATVSGYWKATGKDRTVMDPR
TRQLVGMRLTLVFYRNRAPNGIKTTWIMHEFRLECPNIPHKEDWVLCRVFNKGRDSSLQD
NNYYNNDNQTRLEVNDAPDLNYNQLPPLLSSPPHNHQHEKMKIQVCDQWEQLMKQPSR
TTGHPYHHCHHTIACGWEQMMIGSLSSPSSHGPDHESFAKFALPSTITTVSTSVVII
RIMRRFCCHH*

>G519 (85..894)

CACAAAGATCCTCCGATTCCAAGGTTTATAAAACTCAAATCGAATCTTATCCACAAGA
AAACAACAAGGTACTTTTCCAAAATGAAGCGGAGTTGAATTTGCCGGCGGGATTCCGA
TTTCATCCGACGACGAAGAGCTTGTCAAGTCTATCTTTGCCGGAGATGTGCGTCAGAA
CCGATTAAACGTTCCGGTTATCGCAGAGATTGACTTGTACAAATTCAATCCATGGGAGCTT
CCAGAAATGGCGTTGTACGGTGAGAAAGAATGGTACTTCTCTCGCATAGAGACCGGAAA
TACCCAAACGGGTCGAGACCAAACCGGGCAGCTGGAACCGGTTATTGGAAGCGACTGGA
CTGTATAAACCGATCGGAAAACCGAAGACGTTAGGGATTAAAGAAAGCACTCGTCTTCTAC
GCAGAAAAGCTCCGAAAAGGGATTAAAACGAATTGGATTATGCACGAGTATCGTCTCGCT
AATGTGCGATCGATCTGCTTCTACCAACAAGAAGAACAACTTAAGACTTGATGATTGGGTT
TTGTGTGCGATATACAATAAGAAAGGAACAATGGAGAAGTATTTACCGCGGCGGCTGAG
AAACCGACGGAAGATGAGTACGTGCGACTCAAGATGCTCAAGTCACGTGATTTACCG
GACGTACGTGTTCTGATAACTGGGAGGTTGAGAGTGAGCCCAATGGATTAATCTGGAA
GACGCGTTAGAGGCATTTAATGATGACACGTCCATGTTTAGTTCCATTGGTTTGTGCAA
AATGACGCCTTTGTTCTCAGTTTCAGTACCAGTCCTCCGATTTCTGTCGATTCTGTTTCTAG
GACCCGTTTCGAGCAGAAACCGTTCTTGAATTGGAATTTTGTCTCCTCAAGGGTAAAAATAA
TCGGCAAAAAGTTGAAGCTTTTCAGAGTCTTCGATCACCGGCATTGTGTGCGATCCTGAC
CCGGAGACCAAGTCGGGTCATACGATTACATAATCGGGTTATTGAGATTTCCACATTTGG
ATTTCCGAGACTAACCAACTTAACGGATTCTGGGGTAATTGGGGGGTTTTCACAGGTGA
ATCACACTGAGTCAGCAAGTTTCGATTTTGTGTTTGTGTTTGTAAATGATTGATTAAATG
TCTAAAGATATCAGCAAGTAGATTGAGAAGAACTGTAAGCAATTGTGACCACCCGTTA
TGAATCATAAATATATCAATGAAGCATGAGCTTATTTTTTTTTTAAAAAAA

>G519 Amino Acid Sequence (conserved domain in AA coordinates: 11-104)
MKAEELNLPAGFRFHTDEELVKFYLCRRCASEPINVPVIAEIDLYKFNPWELPEMALYGE
KEWYFFSHRDRKYPNGSRPNRAAGTGYWKATGADKPIGPKTLGIKKALVYAGKAPKGI
KTNWIMHEYRLANVDRSASTNKKNNLRLLDDWVLCRIYNKKGTMKYLPAAEKPTKEMST
SDSRCSHVISPDVTCSDNWEVESEPKWINLEDALEAFNDDTSMFSSIGLLQNDAFVPOF
QYQSSDFVDSFQDPFEQKPFNLWNFAPQG*

>G561 (86..1168)

AATTTGTTTTTTTTCTTTTGTGGGTTCAATTGGAATTGTTTCCCTGAGACTCAAGTTA
CTGTGTCACTACTCTGCAATTGAGCAATGGGTAGCAACGAAGAAGGAAACCCACTAACAA
CTCTGATAAGCCATCGCAAGCTGCTGCTCCTGAGCAGAGTAATGTTTCATGTGTATCATCA
TGACTGGGCTGCTATGCAGGCATATTATGGGCCTAGAGTTGGTATACCTCAATATTACAA
CTCAAATTTGGCGCCTGGTCATGCTCCACCGCCTTATATGTGGGCGTCTCCATCGCCAAT
GATGGCTCCTTATGGGAGCATTATCCACCATTTTGCCCTCCTGGTGGAGTTTATGCTCA
TCCTGGTGTTCAAATGGGCTCACAAACCACAAGGTCCTGTTTCTCAATCAGCATCTGGAGT
TACAACCCCTTTGACCATTGATGCACCAGCTAATTCAGCTGGAAACTCAGATCATGGGTT
CATGAAAAGCTGAAAGAGTTTCGATGGACTTGCAATGTCAATAAGCAATAACAAAGTTGG

GAGTGCTGAACATAGCAGCAGTGAACATAGGAGTTCTCAGAGCTCCGAGAATGATGGCTC
TAGCAATGGTAGTGGTAATAACAACCTGGGGGAGAACAACTAGGAGGAAAAGAAGGCA
ACAAAGATCACCAAGCACTGGTGAAAGACCCCTCATCTCAAAACAGTCTGCCTCTTAGAGG
TGAAAATGAGAAACCCGATGTGACTATGGGGACTCCTGTTATGCCCACAGCAATGAGTTT
CCAAAACCTCTGCTGGCATGAACGGTGTGCCACAGCCATGGAATGAAAAAGAGGTTAAACG
AGAGAAGAGAAAACAGTCAAACCGAGAATCTGCTAGGAGGTCAAGACTGAGGAAGCAGGC
TGAAACAGAACAACTATCTGTCAAAGTTGACGCATTAGTAGCTGAGAACATGTCTCTGAG
GTCTAAACTAGGCCAGCTAAACAATGAGTCTGAGAACTACGGCTGGAGAACGAAGCTAT
ATTGGATCAACTGAAAGCGCAAGCAACAGGGAAAACAGAGAACCTGATCTCTCGAGTTGA
TAAGAACAACCTCTGTATCAGGTAGCAAACTGTGCAGCATCAACTGTTAAATGCAAGTCC
GATAACCGATCCTGTGCGGGCTAGCTGACCGTGGCCGCAACATGAGAACCCGATATTTT
TTCCTTTGGGTTGTGATTGTAACCTAAAAGGAGACTTTTTGTTTTATTCTTAGATTTGT
AGCTCTCTGCATAGTGAGCATAAATGATGTAATATGGTTTAAGAGATTCCGGTGTCTCT
GGTGTGTGCTGCAACCACATAATTGGTGATAGATAGGTTTAGTTATATAAGCAAATGTAT
TAGAGATAAGGGGAGACATATTTGATGGTCTTT

>G561 Amino Acid Sequence (domain in AA coordinates: 248-308)
MGSNEEGNPTMNSDKPSQAAPEQSNVHVYHHDWAAMQAYYGPRVGIPQYYNSNLAPGHA
PPPYMWASPSMMPYGPYPFPFPGGVYAHPGVQMGSPQGPVSQSASGVTTPLTIDA
PANSAGNSDHGFMKKLKEFDGLAMSISNNKVGSAEHSSEHRSSQSSSENDGSSNGSDGNT
TGGEQSRKRQRQSPSTGERPSSQNSLPLRGENEKPDVTMGTPVMPTAMSFQNSAGMNG
VPQPNNEKEVKREKRQSNRESARRSRLRKAETEQLSVKVDALVAENMSLRSLKGLQNN
ESEKRLLENBAILDQLKAQATGKTENLISRVDKNNSVSGSKTVQHQLLNASPIITDPVAAS
*

>G590 (102..1223)

TCGACAGACACTCTCCCTCTCTCCATGCCCATAAATCTCAAAGACTGTTTTAAAAA
AATGTTTTAGCTTTAACTGCTTTTTTTTGTGTTGGTGTAAATGATATCACAGAGAGAAG
AAAGAGAAGAGAAGAAGCAGAGAGTGATGGGAGATAAGAAATTGATTTTCATCTTCTTCT
CTTCTCGGTTTACGATACTCGTATCAATCATCATCTTCATCATCCTCCGTCTTCTTCCG
ACGAAATCTCTCAGTTTCTCCGGCATATTTTCGACCGTTCTTCTCCTTTACCTTCTTACT
ACTCCCCGGCGACGACTACAACGACGGCGTCTTTGATTGGTGTGCACGGGAGCGGTGACC
CACATGCAGATAAATCGAGAAGTCTCGTTTTCTCATCATCCACCGTCAGATTCTGTGCTTA
TGTCGAAACGTGTGCGAGATTTCTCTGAGGTTTTAATCGGCGGAGGATCAGGCTCAGCCG
CCGCGTGTTTTGGTTTCTCCGGTGGTGAATAATAACAACGTTCAAGGAAATAGCTCTG
GGACTCGAGTATCGTCTTCTCCGTTGGAGCTAGTGGCAACGAGACAGATGAGTATGACT
GTGAAAGCGAGGAAGGAGGAGAAGCTGTAGTTGATGAAGCTCCCTCTTCCAAGTCAGGTC
CTTCTTCTCGTAGTTTCATCTAAAGATGCAGAGCTGCTGAAGTTTCATAATCTCTCTGAGA
AGAGGAGGAGAAGTAGAATTAATGAAAAATGAAAGCTTTACAAAGTCTCATCCCTAATT
CAATAAGACGGGATAAGGCTTCAATGCTTGATGAAGCCATTGAGTATCTGAAACAGCTTC
AGCTCCAAGTTTCAGATGTTGACTATGAGAAATGGAATAAACTTGCATCCTTTGTGTTTAC
CTGGAACCTACATTACACCCATTGCAACTCTCTCAGATTTCGACCCCTGAAGCAACCAATG
ATCCTCTGCTTAATCATACCAATCAGTTTGCTTCGACTTCTAATGCACCGGAAATGATCA
ATACTGTGGCTTCTTCATACGCTTTGGAACCTTCTATTTCGCAGTCACTTTGGACCTTTCC
CTCTCCTTACTTCACCCGTGGAGATGAGTCGGGAAGGTGGGTAACTCATCAAGGTTGA
ACATTGGTCATTCCAACGCAACATAACCGGGGAACAAGCTCTGTTTGTATGGACAACCTG
ACCTAAAAGATCGAATTACTTGAACAGTGTCCCACTTCGGGATCTCTATGTGTTCTTGT
TTCTTAGAACGCAAGCCATAAAGCTGTCTGAC

>G590 Amino Acid Sequence (domain in AA coordinates: 202-254)
MISQREERBEKKQRYMGDKKLISSSSSSVYDTRINHHHLHPPSSSDEISQFLRHIFDRS
SPLPSYSPATTTTTASLIGVHSGDPHADNSRSLVSHHPPSDSVLMSKRVGDFSEVLIG
GGSGSAAACFGFGGNNNNVQGNSSGTRVSSSVGASGNETDEYDCESEEGGEAVVDEA
PSSKSGPSSRSSSKRCRAAEVHNLSEKRRRSRINEKMKALQSLIPNSNKTDKASMLDEAI
EYLKQLQLQVQMLTMRNGINLHPLCLPGTTLHPLQLSQIRPPEATNDPLLHNHTNQFASTS
NAPEMINTVASSYALEPSIRSHFGPFPLLTSPVEMSREGGLTHPRLNIGHSNANITGEQA
LFDGQPDLDKDRIT*

>G818 (65..1060)

GTATTCTTACAATAAACGACCAAAAAGTTAATACAAGAAATAGAAACGGTGTAGGAAGC
TACTATGACGGCAATTCCAAACGTCGTCGATATTGAATCTTCTTCTCTCGCTTTGTCA

AGAGACGGCAACGGAGACCGTCACCGTTGAAAGAGGCTCGTCTGATTTCATCTTCAAAGCC
AGACGACGTCGTTTTACTAATCAAGGAAGAGGAGGATGACGCCGTTAACTTGTCACTTGG
TTTTTGGAAATTGCACGAGATAGGTTTAATAACACCGTTCTTGAGAAAGACGTTTGAGAT
CGTCGATGACAAAGTAACAGACCCCGTTGTATCATGGAGCCCGACCCGTAAAAGCTTTAT
CATTTGGGATTCTTACGAGTTCTCAGAGAATCTACTTCCCAAATACTTCAAGCACAAGAA
CTTCTCCAGTTTTATTTCGTCAGCTTAACTCTTACGGTTTTAAAAAGGTCGATTTCAGATAG
GTGGGAATTTGCTAACGAAGGGTTTCAAGGAGGGAAGAAACATTTGCTTAAGAACATCAA
GAGGAGAAGCAAAACACTAAATGTTGTAACAAGGAAGCGAGTACCACCACGACAGAGAC
TGAGGTTGAGTCATTGAAGGAGGAACAGAGTCCAATGAGATTGGAGATGTTGAAGCTGAA
ACAACAACAAGAAGATCTCAACATCAGATGGTCACTGTGCAGGAGAAGATCCACGGAGT
TGATACCGAACACAGCATATGCTTAGTTTCTTGCAGAGTTGGCTAAAGATCAAAGATT
TGTAGAGAGACTGGTGAAGAAGAGAAAGATGAAAATACAGAGAGAGCTAGAAGCAGCTGA
ATTCTGTGAAGAAGCTCAAGTTGCTTCAGGATCAAGAACTCAAAAGAACTTGTAGATGT
AGAAAGAGAATTTATGGCCATGGCTGCAACAGAACACAATCCCGAGCCTGACATTTTGGT
GAACAATCAAAGCGGGAATACGAGATGTCAGCTTAACTCAGAGGACCTACTTGTGACGG
TGGCTCAATGGATGTAAATGGGAGGATAGAGATAGAGTAGAGCAAAACCGGTAACATAGC
AATAGAGAAGGTACCAAATCCCAAGGCTTGAGATCCGAAT

>G818 Amino Acid Sequence (domain in AA coordinates: 70-162)
MTAIPNVVDIESSSSSLCQETATETVTVERGSSDSSSKPDDVLLIKEEDDAVNLSLGF
WKLHEIGLITPFLRKTFEIVDDKVTDPVVSWSPTRKSFIIWDSYEFSENLLPKYFKHKNF
SSFIRQLNSYGFKKVSDRWEFANEGFQGGKKHLLKNIKRRSKNTKCCNKEASTTTTETE
VESLKEEQSPMRLEMLKLKQQEESQHQMVTVOEKIHGVDTEQQHMLSFFAKLAKDQRFV
ERLVKKRKMKIQRELEAAEFVKKLKLQDQETQKNLLDVEREFMAMAATEHNPEPDILVN
NQSGNTRCQLNSEDLLVDGGSMDVNGRIEIE*

>G849 (218..2077)

AACTCGAGAATCTTTCATTTCTTTTAAATCTTAGAATCTCGAGTTTTTGTATAAATCGAT
TCTAATTTTTCTTTTGTACATTGTTTTATATATACATAAAACACACAAATCGGGTATGGG
GGAATTTGGGTTTTAAGATAGCGTGATCTGTAATAATAAGTGGTTCGCGATCGTGATCAA
GAAACTGGTGGCTGATAGTGATATGCATATTTGAGAGATGGTGTTCAGAGAAAGTTAGA
TTGCCTTTCCGTGGGATTTGATTTTCCCAACATTCCCAGAGCTCCTCGTTCATGCAGGAG
GAAGGTTCTAAACAAGAGGATTGATCATGATGATAACACTCAGATCTGTGCAATTGA
CTTACTAGCTTTGGCTGGAAAGATTCTACAGGAAAGCGAGAGTTCCTCTGCGTCTTCTAA
TGCATTTGAAGAAATTAAGCAAGAGAAAGTAGAAATTCGAAGACTATTAAATCTGAGTC
TTCTGACCAAGGAAACTCTGTGTCAAAGCCTACTTATGATATCTCTACTGAGAAGTGTGT
GGTGAACAGTTGTTTTTTCATTTCCGGATAGTGACGGCGTTTTGGAGCGGACTCCGATGTC
TGATTACAAGAAGATTTCATGGTTTGATGGATGTAGGGTGTGAAAACAAGAATGTAAATAA
TGGGTTTCGAGCAAGGAGAAGCAACCGATCGCGTGGGTGATGGAGGCTTAGTCACTGATAC
TTGCAACTTAGAGGATGCAACTGCGTTAGGTCTGCAGTTTCCGAAATCAGTCTGTGTGGG
TGGTGAATTTAAATCACCATCCACCTTGGATATGACCCCTAATGGTTCCATATGCTAGACA
TGGGAACCATACTAACCCTAGGTAGAAAAGATGATGATGAAAAATTCTATAGTTACCATAA
ACTTAGCAATAAATTTAAGTCGTATAGGTCTCCAACAATTCGAAGAATAAGAAAGTCCAT
GTCGTCCAAATACTGGAACAAGTTCCAAAAGATTTTGGATAACAGTAGAGCTGATGTGGG
TGTGAAGACTCTTTATCGCAAAAGAAAATCATGTTATGGTTACAACGCATGGCAGCGTGA
GATCATTTATAAGAGAAGAAGATCACCTGACAGAAGCTCGGTGTAACCTTCTGATGGAGG
ACTCAGTAGTGGAAGTGTTTCCAAGTTACCCAAGAAGGGAGATACAGTAAAGCTAAGCAT
TAAGTCCTTTAGGATTCAGAGCTTTTTATTGAAGTTCCAGAACTGCAACAGTAGGATC
ACTAAAGAGGACTGTGATGGAGGCTGTCTAGTGTCTTACTCAGCGGAGGAATACGTGTTGG
GGTGTAAATGCATGGGAAGAAGGTTAGAGATGAAAGGAAAACCTCTGTCCAGACTGGGAT
CTCATGTGATGAAAATCTAGACAACCTTGGGTTACCTTGGAGCCTAGTCCCAGCAAAGT
TCCCCTACCTTTGTGTTCTGAAGATCCTGCTGTGCAACCGACCTACAAGTTTGTCTGA
ACGGTCTGCGGCGTCTCCTATGCTAGATTCTGGAATTCACATGCAGATGACGTGATTGA
TTCAAGAAATATTGTGGACAGTAACCTCGAATTAGTTCCATATCAGGGTGACATATCTGT
TGATGAACCTTCATCAGATTCAAAGAGCTTGTCCCACTTCAGAGTTGGAAGTCAAGGC
GCTTGCCATAGTTCCGTTGAACCAGAAACCTAAGCGTACTGAGCTAGCCCAGAGGAGAAC
TAGGAGACCCTTCTCTGTGACAGAGGTAGAAGCTTGTGACAAGCAGTTGAGGAACCTCGG
GACTGGAAGATGGCGTGATGTAAATTCGCTGCTTTCGAGGATGCAGATCATCGGACTTA
CGTGGAATTGAAGGACAAATGGAAGACGCTAGTTCACACAGCAAGTATATCCCCACAGCA

ACGAAGAGGAGAGCCGGTGCCACAAGAACTGCTAGACAGAGTCTTGAGGGCATAACGGGTA
TTGGTTCGCAGACCAAGGAAAACATCAGGCGAGAGGAGCGTCCAAAGATCCAGACATGAA
CAGAGGTGGAGCTTTTGAATCAGGTGTTTCAGTGTAAGGAGGAGGTACGCATTGGTGGG
TGGGTGTACAGAAGCAAACAACACAATAAATGGACAACCTCAATTTCTGCAAAGTTTAATT
GTCTTTATTTCTCGTTTTTTTTTTTTTCTCTCTACATACACTTTTTTTTTTCT
>G849 Amino Acid Sequence (domain in AA coordinates: 324-413, 504-583)
MVFKRKLDCLSVGFDFPNIPRAPRSCRRKVLNKRIDHDDNTQICAILDLALAGKILQES
ESSSASSNAFEEIKQEKVENCKTIKSESSDQGNSVSKPTYDISTEKCVVNSCFSPDSDG
VLERTPMSDYKKIHGLMDVGCENKNVNNNGFEQGEATDRVGDGGLVTDTCNLEDATALGLQ
FPKSVCVGGDLKSPSTLDMTPNGSYARHGNHTNLGRKDDDEKFYSYHKLSNKFYSYRSP
IRRIKSMSSKYWKQVPKDFGYSRADVGVKTLYRKRKSCYGYNAWQRELIYKRRRSPDRS
SVVTSDDGLSSGSVSKLPKKGDTVKLSIKSFRIPELFIEVPETATVGSILKRTVMEAVSVL
LSGGIRVGVLMHGKVKVDERKTLSTGTISCDENLDNLGFTLEPSPSKVPLPLCSEDPAVP
TDPTLSLERSAASPMLDGIPHADDVIDSRNIVDSNLELVPIYQGDIVDEPSSDSKELVP
LPELEVKALAIIVPLNQPKRTELAQRRTRRPFSVTEVEALVQAVEELGTGRWRDVKLRAF
EDADHRTYVDLKDWKTLVHTASISPPQRRGEPVPQELLDRVLRAYGYWSQHQQKHQARG
ASKDPDMNRGGAFESGVSV*

>G892 (21..1004)
TATAACAATTCTCTCAACAATGTCTATTGAGTCAGCCAATAACACGGACCGATAGTGCAC
CCAATGGAGCATTTAGGACTTTTGGTCTCTACTGGTGTCTACCATTGTGATCGTATGGTCA
GAATTGCATCCTCTAACCCATCAGAGATCGCCTGTCTCGATGTTTGAGGCAATTTGTCTG
TTGAGATTGAAACGAGACAACGGCCTCGGTTTACTTTCAACCATGCTACTCCGCTTTTG
ATGCTTCTCTGAGGCTCGTCTTCTCGAAGCTCTCTCGCTCATGTTTGAGCCTGCAACCA
TAGGTAGGTTTGGTGCAGACCCATTTCTTAGGGCAAGATCCAGAAACATCTTGAACCTG
AATCAAGACCCCGACCGCAACATCGAAGACGACACAGCCTTGACAATGTTAACAATGGTG
GTTTACCTCTACCAAGAAGAACATATGTTATTCTCCGGCCCAATAATCCGACTAGTCCAC
TCGGAAACATAATTGCGCCACCAATCAAGCACCACCACGGCATGTGAACCTCACATGATT
ACTTTACTGGAGCATCAAGCTTAGAGCAGCTGATTGAACAACCTAACACAAGACGATAGGC
CTGGACCACCACCTGCGTCAGAACCCACCATTAAATCCCTACCATCTGTGAAAATAACAC
CACAACATCTAACTAACGACATGTCCCAATGCACAGTGTGCATGGAAGAATTCATTGTTG
GTGGGGACGCAACGGAATTACCATGTAAACATATTTACCATAAAGATTGTATAGTCCCGT
GGCTTAGGCTTAAACAATTCTTGCCCTATCTGCCGCCGTGACCTGCCACTTGTCAACACCG
TTGCTGAATCTCGAGAAAGGAGCAATCCTATTAGACAAGACATGCCTGAAAGAAGGCGTC
CAAGGTGGATGCAACTCGGTAACATTTGGCCATTTAGAGCAAGATAACCAAGGGTTAGTC
CAGAAGAAACAGCAAACCAAGAAATCCTCGAGATAACAGGAGCTAACTCTGAATATTCATG
GGAAATAAAATCGTGACTATCTATATGTATAGACTCTATGAGACATTGTCTATTTGAAT
GTGCATGTATATCTCAGAAATAAATCAAGCGAAACATATTTAACGACTAAAAAAA

>G892 Amino Acid Sequence (domain in AA coordinates: 177-270)
MSLSQPITRTDSAPNGAFRTFGLYWCYHCDRMVRIASSNPSEIACPRCLRQFVVEIETRO
RPRFTFNHATPPFDASPEARLLEALSMLFEPATIGRFADPFLRARSNILEPESRPRPQ
HRRRHSLDNVNNGGLPLPRRTYVILRPNNPTSPLGNI IAPPNQAPPRHVNSHDYFTGASS
LEQLIEQLTQDDRPGPPASEPTINSLPSVKITPQHLTNDMSQCTVCMEEFIVGGDATEL
PCKHIYHKDCIVPWLRLNNSCPICRRDLPLVNTVAESRERSNP I RQDMPERRRPRWMQLG
NIWPFRRARYQRVSP EETANQNPRDNRS*

>G961 (1..1200)
ATGTCAAAATCTATGAGCATATCAGTGAACGGACAATCTCAAGTGCCTCCTGGGTTTAGG
TTTCATCCGACCGAGGAAGAGCTGTTGCAGTATTATCTCCGGAAGAAAGTTAATAGCATC
GAGATCGATCTTGATGTCTATTCGCGACGTTGATCTCAACAAGCTCGAGCCTTGGGACATT
CAAGAGATGTGTAATAAGGAACAACGCCACAAAACGACTGGTATTTCTTTAGCCACAAG
GACAAAAATATCCGACGGGAACGAGAACTAACAGAGCCACTGCGGCTGGATTTTGAAA
GCAACTGGCCGCGACAAGATCATATATAGCAATGGCCGTAGAATTGGGATGAGAAAGACT
CTTGTTTTCTACAAAGGCCGAGCTCCTCACGGCCAAAAATCTGATTGGATCATGCATGAA
TATAGACTCGATGACAACATTATTTCCCCCGAGGATGTACCGTTCATGAGGTCTGTGAGT
ATTATAGGGGAAGCATCACAAGACGAAGGATGGGTGGTGTGTCGTATTTTCAAGAAGAAG
AATCTTCACAAAACCTAAACAGTCCCGTCGGAGGAGCTTCCCTGAGCGGCGCGGAGAT
ACGCCGAAGACGACATCATCTCAGATCTTCAACGAGGATACTCTCGACCAATTTCTTGAA
CTTATGGGGAGATCTTGTAAGAAGAGCTAAATCTTGACCCTTTCATGAACTCCCAAAC

CTCGAAAGCCCTAACAGTCAGGCAATCAACAACCTGCCACGTAAGCTCTCCCGACACTAAT
 CATAATATCCACGTGAGCAACGTGGTCGACACTAGCTTTGTTACTAGCTGGGCGGCTTTA
 GACCGCCTCGTGGCCTCGCAGCTTAACGGACCCACATCATATTCAATTACAGCCGTCAAT
 GAGAGCCACGTGGGCCATGATCATCTCGCTTTGCCTTCCGTCCGATCTCCGTACCCAGC
 CTAAACCGGTCCGCTTCGTACCACGCCGTTTAAACACAGGAATATACACCGGAGATGGAG
 CTATGGAATACGACGACGTCTCTATCGTCATCGCCTGGCCCATTTTGTACGTGTCTG
 AATGTTTTGCTGCTTTGTTGTCTCTCTTCGTCTGCAGCTTCAGTTCTGGCCGTTCCAACCA
 TGGCAGAGGCAGGTTTCATTTTCATCTTCATCGCCTCAGATGCAGATCTCTCTCCATTGA
 >G961 Amino Acid Sequence (conserved domain in AA coordinates: 15-140)
 MSKSMSISVNGSQVPPGFRFHPTBEEELLQYYLRKKVNSIEDLDVIRDVDLNKLEPWDI
 QEMCKIGITTPQNDWYFFSHKDKKYPTGTRTNRATAAGFWKATGRDKIIYSNGRRIGMRKT
 LVFYKGRAPHGQKSDWIMHEYRLDDNIISPEDVTVEHVSIIGEASQDEGWVVCRIFFKKK
 NLHKTLSNPVGGASLSGGGDTPKTTSSQIFNEDTLDQFLELMGRSCKEELNLDPFMKLPN
 LESPNSQAINNCHVSSPDTNHNHVSINVVDTSFVTSWAALDRLVASQLNGPTSYSITAVN
 ESHVGHDLALPSVRSPPSLNRSASYHAGLTQEYTPEMELWNTTSSLSSSPGPFCHVS
 NVLLLVCLLRLQLQFWFPQWQRQVHFDLSSPQMQLSLH*

>G1465 (163..1125)
 TATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTTCATTTGGAGAGGACACGC
 TGACAAGCTGACTCTAGCTTATCTGGTACCGTCGACCTCATTCTTGCCTTTGATCTTTCT
 TTCTCTAGATCCCATATTTTTCTTGATCAATTTAGTTTCATTATGGAGGAAGATGCAGCT
 TTTGATCTACTCAAAGCCGAACCTCTTAAACGCAGAAGACGATGCAATAATCTCACGTTAT
 CTGAAGCGTATGGTTCGTCACCGGAGACTCATGGCCTGATCATTTCATCGAAGACGCAGAC
 GTGTTCAACAAGAATCCAAATGTGGAGTTCGATGCTGAGAGCCCTAGCTTCGTGATAGTT
 AAACCTCGAACAGAGGCTTGTGGTAAACCGATGGATGTGAACTGGTTGCTGGAGGATC
 ATGGGTCTGTGATAAAACCGATAAAATCGACGGAGACTGTGAAGATTCAAGGGTTCAAGAAG
 ATTCTCAAGTTCCTGCCATAAGAGGAAACCTAGAGGATACAAGAGAAGTTGGGTAATGGAA
 GAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAGATCATGTGATTGCAAGATTTCGG
 TTTATGTTTGAAGCTGAAATCAGTTTCTTGCTAGCCAAGCATTCTTACACTACATCAGAA
 TCACCTTCCTCGAAATGAGCTGTTGCCAGCTTACGGATTCTTTTCATCAGATAAGCAATTG
 GAGGATGTATCTTATCCGGTGACGATAATGACTTCTGAAGGAAACGATTGGCCTAGCTAC
 GTTACCAACAATGTGTATTGTCTGCATCCATTGGAGCTCGTTGATCTTCAAGATCGGATG
 TTTAATGATTACGGAACCTGCATCTTCGCTAACAAGACTTGTGGTAAACCGATAGATGC
 ATTAATGGTGGTTACTGGAAAATTTTGCACCGTGATAGGCTGATCAAGTCAAAGTCCGGG
 ATAGTTATTGGTTTCAAGAAGGTGTTTAAGTTTCATGAAACGGAGAAAGAAAGATACTTC
 TGTGGTGGAGAAGATGTGAAGGTAACCTGGACTCTAGAAGAGTATAGGCTTAGCGTGAAG
 CAGAATAAATCTTGTGCGTTATCAAGTTTACTTATGATAACTAAGAATCTTTTCTTTGG
 ATTTTATGATCATCTTAGTATCGCGACCGCTCTAGACAGGCCTCGTACCGGATCCTCTAG
 CTAGAGCTTTTCGTTTCGTATCATCGGTTTCGACAACG
 >G1465 Amino Acid Sequence (conserved domain in AA coordinates: 242-306)
 MEEDAAFDLLKAELLNAEDDAIISRYLKRNVNGDSWPDHFIEDADVFNKNPNVEFDAES
 PSFVIVKPRTEACGKTDGCETGCRIMGRDKPIKSTETVKIQGFKKILKFLKRKPRGYK
 RSWVMEEYRLTNLNLWKQDHVICKIRFMFEAEISFLAKHFYTTSESILPRNELLPAYGFL
 SSDKQLEDVSYPTVITMTSEGNWPSYVTNNVYCLHPLLELDLQDRMFNDYGTICIFANKTC
 GKTDRCINGGYWKILHRDLIKSKSGIVIGFKKVFKEHETEKERYFCGGEDVKVWTWLEE
 YRLSVKQNKFLCVIKFTYDN*

>G425 (45..1196)
 GAAAACAGTCTTCTCTTCTCCGATCCCAAAAACGCAGGAAAACAATGTGCTTTAACAGCTCCC
 ACCTCCTTCCCTCCAACAAGAAGACCTTCTCTCCGACACTTCACCGATCAATCACAGCAACCTC
 CGCCGACGCGTCACTTCTCTGAAACACCTTCGCTTGTACCGCCAGTTTCTCAACCTCCCTA
 CCACCTTACCATTGCGGATCCGATCTCGCTCCTCCGACCGCAACGGAGACAATTCCGTT
 GCTGATACAAACCCACGCTGGCTCTCTTTTCATTTCGGAGATGCAAAATACTGGAGAAGTACG
 TTCTGAAAGTTATCGACGGAGTCAACGCCGATGGTGAAACGATACTCGGCGTTGTAGGAGGT
 GAAGATTGGCGGAGTGCTAGCTATAAGGCGGCGATTTTAAAGACATCCGATGTACGAGCAGC
 TTCTTGGCGCTCACGTGGCTTGCCTTAGGGTTGCGACTCCCGTTGACCAGATTCCGAGGATC
 GATGCTCAGCTCAGTCAGTTGCATACCGTCGCCGCGAAATACTCCACTCTTGGTGTGGTTGTT
 GACAACAAGGAACCTTGATCATTTTCATGTACATTATGTTGTCTTGTATGTTTCATTTAAAGAACA
 ACTCCAACACCACGTTTGTGTCCATGCAATGGAAGCCATTACGGCTTGTGGGAGATTGAACA

>G2069 (1..1026)

ATGGAAGGAGGAGGAAGAGGACCAAATCAAACGATTCTCAGTGAAATAGAACATATGCCT
GAAGCTCCACGTCACGTATCTCTCATCACCCTCGAGCTCGCTCTGAAACCTTCTTCTCC
GGCGAATCAATCGACGATCTCCTCTTATTCGATCCTTCCGATATCGATTCTCTTCTCTA
GACTTCTCTCAACGCTCCACCACCACCAACAATCACAACAACAACCGCAAGCTTCTCCC
ATGTCCGTTGATTCCGAAGAAACCTCATCGAACGGTGTGTCTCTCTAATTCTCTTCTCT
CCAAAACCCGAAGCTAGATTTCGGTCGCCATGTTCTGAGCTTCTCGGTTGATTCCGATTTC
TTCGATGATTTGGGTGTTACTGAGGAGAAGTTTATAGCTACAAGTTCAGGAGAGAAGAAG
AAAGGGAAATCATCATCATAGCAGGAGTAATTCTATGGATGGAGAGATGAGTTCGGCGTCG
TTTAATATCGAATCGATTTTAGCTTCTGTGAGTGGTAAAGATAGTGGGAAGAAGAATATG
GGTATGGGTGGTGATAGACTTGCTGAGCTTGCTTTGCTTGATCCTAAAAGAGCTAAAAGG
ATTTTAGCGAATAGACAATCTGCGGCGAGGTCGAAAGAGAGGAAGATTAGGTATACCTGGT
GAGTTAGAGAGGAAGTTTCAGACACTTCAGAATGAAGCTACTACATTGTCTGCTCAAGTC
ACTATGTTACAGAGAGGAACATCAGAGCTGAACACTGAAAATAAACACCTCAAAATGCGG
CTTCAAGCTTTAGAGCAACAAGCTGAACCTAGGGATGCTTTGAATGAAGCGCTGCGGGAT
GAACTGAACCGACTTAAGGTGGTAGCTGGAGAAATTCCTCAGGGGAATGGAATTCCTTAC
AACCCTGCTCAATTCTCATCTCAGCAATCGGCAATGAATCAGTTTGGGAACAAAACGAAC
CAACAGATGAGTACAAACGGGCAGCCATCGCTCCCAAGCTACATGGATTTCACCAAGAGA
GGCTGA

>G2069 Amino Acid Sequence (domain in AA coordinates: TBD)

MEGGGRGPNQILSEIEHMPAPRQRISHHRRARSETFFSGESIDDLILFDPSDIDFSSL
DFLNAPPPPPQSQSQPQASPMVDSEETSSNGVPPNSLPPKPEARFGRHVRFSFVDSDF
FDDLGVTEEFKFIATSSGEKKKGNNHHSRSNSMDGEMSSASFNIESILASVSGKDSGKKNM
GMGGDLRLAELALLDPKRAKRIILANRQSAARSKERKIRYTGELERKVQTLQNEATTLQAQV
TMLQRGTSELNNTENKHLKMLRLQALEQQAELRDALNEALRDELNLKLVVAGEI PQGNNGNSY
NRAQFSSQQSAMNQFNGKTNQOMSTNGQPSLPSYMDFTKRG*

>G1852 (55..1857)

CATCTGATCTGCTCTCGAAGACGAAAGCTTCGAGTACTGGTTGAAGCTAAAGCTATGGGA
CACGTGAATCTACCTGCATCAAAGCGTGGTAACCTCGTCAATGGCGTCTCCTCGACATC
GTAACCGCTGCTTCTTCGGTATCGTACTTCTCTTCTTCTCATCCTTTTATTCACCTCCTT
GGTGATTCCATGGCGGCTTCTGGTCGGCAAACGCTGCTTCTCTTACGGCGTCAGATCCG
AGGCAACGGCAGCGATTAGTGACTTTGGTTGAAGCTGGTCAGCATTTCGAACCGATCGAG
TATTGTCTCTGCGGAAGCTGTTCATATGCCTTTGTGAGGATCCGAGAAGGAATAGTCAG
CTTAGTAGAGAGATGAATTTCTATAGGGAGAGACATTGTCTTTGCTGAGGAGACTCCG
CTCTGTTTGATTCTCCGCTTCTGGTTATAAAATTCCTGTTCGGTGGCCTGAGAGTCTT
CACAAGATTTGGCATGCAAAATGCCATATAACAAAATTCTGACCGGAAAGGTTCATCAA
GGATGGATGAAAAGGGAAGGGGAATACTTTACTTTCCAGGCGGTGGCAGATGTTTCCT
GGCGGAGCTGGCCAAATACATTGAAAAGCTTGACAGTATATTCCGCTTAATGGTGGAACT
TTGAGAACTGCTCTTGACATGGGATGCGGGGTAGCTAGTTTGGAGGTACTCTACTATCT
CAAGGCATTCTAGCCCTCTCATTGTCTCCAAGAGATTCACATAAATCTCAAATTCAGTTC
GCTTTGGAAAGAGGAGTGCCCTGCATTTGTTGCCATGCTTGGCACTCGTAGACTCCCTTT
CCTGCATACTCCTTTGACCTGATGCACTGTTCCCGATGTTTGATTCTTTACGGCTTAC
AATGCAACTTACTTCATCGAAGTAGATAGGTTACTGCGCCCTGGAGGATATCTTGTAATC
TCTGGCCACCTGTACAATGGCCTAAACAAGACAAAGAATGGGCTGATCTTCAGGCGGTG
GCTAGAGCTTTGTGCTATGAGCTAATTGCGGTTGATGGAAACACTGTCATCTGGAAGAAG
CCTGTTGGAGATTCTATGTCTACCTAGCCAGAATGAGTTTGGGCTTGAGTTGTGTGATGAG
TCTGTTCCGCCAAGTGATGCATGGTATTTTAAATTGAAGAGGTGTGTTACCAGGCCATCA
TCCGTCAAAGGAGATCACGCTTTGGGAACTATATCCAAGTGGCCGGAGAGGCTTACTAAA
GTTCTCTCTAGGGCCATTGTCTGAAAAACGGATTGGATGTGTTTGAAGCAGATGCAAGG
CGGTGGGCAAGACGCGTTGCTTTATTACAGGGATTCTCTTAAGCTGAAATCTCCA
ACTGTCCGCAATGTCTAGGACATGAACGCAATCTTTCGGAGGCTTTGACAGCAACCCCTGCA
TCTGATCCTGTGTGGGTTATGAATGTCTATCCAGCTCGGAAGCCATTAACTCTTGACGTG
ATTTATGACAGAGGTCTCATCGGTGTTTACCATGATTGGTGTGAACCATTTTCAACATAT
CCCCGCACGTATGATTTTCATCCATGTATCAGGAATTGAATCACTGATAAAACGACAAGAC
TCAAGCAAATCGAGGTGTAGCTTAGATCTAATGGTAGAGATGGACAGAATATTACGT
CCAGAGGAAAGGTTGTGATCCGAGACTCTCTGAGGTGCTAGATAAAGTCGCACGAATG
GCTCATGCTGTAAGATGGTCTTCTTCCATACACGAGAAAGAACCTGAATCCCATGGAAGA

YDMPSSDGTGGYSGWTSSESVQGSNPGGVFTMWNE*

>G761 (521..1549)

GGGGCCGACCGGCCCGCCGGGCAGGTCTAGGTTCAAAAAGGACTCACAAGAGAGAGATAGT
 ATGATTGATAGGGAAGAGAGAGATGAAAGAAAGTAAATATATAATAGATTATTAGG
 ACACGAGTGTCTCTTTTGTGTTGTGCTCTCTCTTTCTCTCTCTCTCGAA
 TGATCATCTTTATATAACCTACTCTCTTTCTCTTTTCCATTCTTTTCATATCATTCTCC
 CTTTCTCTCTCGGGATCTGATCTCTCTTTCCAGTAACCTATTCCCGAGGAGCACTGTCAA
 ATCTTGTCCACTCTTTGATCTTATCTCGATCTCTTTCTCTTTCTAGTCTTGTGTAGTCTT
 CAACTTGTGATGTTATCTATATAGTAATCACGAGAGAGAATCATACAATAGCTGAAACA
 TAAAGCTTTCTTAGAAGCTTTAAAAAGGTCTCATCTGGATTATCCTGTTTAAATTTCTAGA
 GTTTCTTCAGGCAGATTATTAACCGATCAAGAAGACAAAATGAATTCATTTTCCCACGT
 CCCTCCGGGTTTTAGATTTCACCCGACAGATGAAGAACTTGTAGACTACTACCTGAGGAA
 AAAAGTCGCATCGAAGAGAATAGAAATTGATTTTATAAAGGACATTGATCTTTACAAGAT
 TGAGCCATGGGACCTTCAAGAGTTGTGCAAAATTGGGCATGAAGAGCAGAGTGATTGGTA
 CTTCTTTAGCCATAAAGACAAGAAGTATCCACAGGGACTCGAACCAATAGAGCAACAAA
 AGCAGGTTTTTGGAAAGCCACCGGAAGAGATAAGGCTATCTATTTGAGGCATAGTCTAAT
 TGGCATGAGGAAAACTTGTGTTTTACAAGGGAAGAGCCCCAAATGGACAAAAGTCTGA
 TTGGATCATGCACGAATACCGCTTAGAAACCGATGAAAACGGAACCTCCTCAGGAAGAAGG
 ATGGGTTGTGTGTAGGGTTTTCAAGAAGAGATTGGCTGCAGTTAGACGAATGGGAGATTA
 CGACTCATCCCTTCACATTGGTACGATGATCAACTTTCTTTTATGGCCTCCGAGCTCGA
 GACAAACGGTCAACGACGGATTCTCCCCAATCATCATCAGCAGCAGCAGCAGCAGCAGCA
 ACAACATATGCCATATGGCCTCAATGCATCTGCTTACGCTCTCAACAACCTTAATTTGCA
 ATGCAAGCAAGAGCTAGAACTACACTACAACCACCTGCAATCAAATATCGCGCATGAGGA
 ACAATTGAATCAAGGAAATCAGAACTTCAGCTCTCTATACATGAACAGCGGCAACGAGCA
 AGTGATGGACCAAGTCAAGACTGGAGAGTTCTCGATAAATTTGTTGCTTCTCAGCTAAG
 CAACGAGGAGGCTGCCACAGCTTCTGCATCTATACAGAATAATGCCAAGGACACAAGCAA
 TGCTGAGTACCAAGTTGATGAAGAAAAAGATCCGAAAAGGGCTTCAGACATGGGAGAAGA
 ATATACTGCTTCTACTTCTTCGAGTTGTGAGATTGATCTATGGAAGTGAGCTGAAAGAGA
 AGACATATAAATGCATATATACATATATATATATACGTACACAGCAACTAATCAAGTG
 TAGATGATGATGATGGTACAGATTTATATTTGCTTTGATTGATTCTTACTACATTATTGA
 ACTTATGTCATATGCATATATACATTGCGTATCTATGCATATTTATACTTGTACTCAATA
 TGATTAACCATATATAAATCTAATCTAAATGTAACCTCCAATATTTTAAATAGACAAT
 TGTCTCTTCTTATTAGAAAAAAA

>G761 Amino Acid Sequence (domain in AA coordinates: 10-156)
 MNSFSHVPPGFRFHPTDEELVDYLRKKVASKRIEIDFIKDIDLYKIEPWLQELCKIGH
 EEQSDWYFFSHKDKKYPTGTRTNRAKAGFWKATGRDKAIYLRHSLIGMRKTLVFYKGRA
 PNGQKSDWIMHEYRLFTDENGTPQEEGWVVCRVFKRLAAVRRMGDYDSSPSHWYDDQLS
 FMASELETNGQRRILPNHHQQQHEHQHMPYGLNASAYALNNPNLQCKQELHYNHLQ
 SNIAHEEQNLQGNQNFSSLYMNSGNEQVMDQVTDWRVLDKFVASQLSNEEAATASASIQN
 NAKDTSNAEYQVDEEKDPKRASDMGEEYTASTSSSCQIDLWK*

>G1056 (10..798)

GCTACATATATGGGTCTATTAGAGGAAACATTGAAGAGCCTATATCTCAGTCATTAACG
 AGGCAGAACTCTCTCTATAGCTTAAAGCTCCATGAGGTTCAAACCCACTTAGGAAGTTCT
 GGAAAACCACTAGGAAGCATGAACCTTGATGAGCTTCTCAAGACTGTCTTGCCACCAGCT
 GAGGAAGGGCTTGTTTCGTGAGGGAAGCTTGACGTTACCTCGAGATCTCAGTAAAAAGACA
 GTTGATGAGGTCTGGAGAGATATCCAACAGGACAAGAATGGAAACGGTACTAGTACTACT
 ACTACTCATAAGCAGCCTACACTCGGTGAAATAACACTTGAGGATTTGTTGTTGAGAGCT
 GGTGTAGTGACTGAGACAGTAGTCCCTCAAGAAAATGTTGTTAACATAGCTTCAAATGGG
 CAATGGGTTGAGTATCATCATCAGCCTCAACAACAAGGGTTTATGACATATCCGGTT
 TGGAGATGCAAGATATGGTGATGATGGGTGGATTATCGGATACACCACAAGCGCCTGGG
 AGGAAAAGAGTAGCTGGAGAGATTGTGGAGAAGACTGTTGAGAGGAGACAGAAGAGGATG
 ATCAAGAACAGAGAATCTGCAGCAGCTTACGAGCTAGGAAACAGGCTTATACACATGAA
 TTAGAGATCAAGGTTTCAAGGTTAGAAGAAGAAAACGAAAACTTCGGAGGCTAAAGGAG
 GTGGAGAAGATCCTACCAAGTGAACCACCACAGATCCTAAGTGAAGCTCCGGCGAACA
 AACTCTGCTTCTCTCTGATCCTAAAGACTCTTCTTTCTTTCTTCTTGTGTTGGTTT
 ATATCAGACCGCTTTGTTCTTTGTATATTGTGTAGACTTTATTGACTTTGAACAGCATGT
 CTTTATAAACATTTCTTGAGTGT

>G1056 Amino Acid Sequence (domain in AA coordinates 183-246)
 MGSIRGNIEEPISQSLTRQNSLYSLKLHEVQTHLGSSGKPLGSMNLDELLKTVLPPEEG
 LVRQGSLLTPRDLSSKKTVDDEVWRDIQQDKNGNGTSTTTTHKQPTLGEITLEDLLLRAGVV
 TETVVPQENVVNIASNGQWVEYHHQPQQQGFMTYPVCEMQDMVMGGLSDTPQAPGRKR
 VAGEIVEKTVERRQKRMIKNRESAARSARKQAYTHELEIKVSRLEBENEKLRLRLKEVEK
 ILPSEPPDPKWKLRRTNSASL*

>G1447 (82..1086)

AAAAACCCTAACCTAATTCTCTCAAGACAACCTCAAAGGTCTCTCCTTTTTTAGGTTTAT
 TATCACTTCCGTATAATCGCCATGTCTTCTCTACCATGGAAAAACCAAATCGAGTCGA
 ATCTTAAGATTCAATTTCTGAGTTTCAACAATCACCGTTCGTTGAAACTGGCTTTCCAACT
 TCTCTGATCGATCTCTTCTTCAAGAATCGCGATCGTCTAAAAAATCTCCATCTAAACGC
 TTCCAACGAATCGAACGCCAGATTCTGAACCGCTCCAAACGCTTCTTCGTTGAGTAATCAA
 GATACGATTTTGAAGAGCCCTCGAGGATTAAACCGTTTGAAGTAAGGTCGAGAAAGTT
 AATTGCGTTAAAGGTAAATCAGCGGCGTGAAGAAGACGCGATTAAAAATAGCGTTTTC
 GGCGGTAGCGGTGAGGTCGTTTTGATGGCGTTTAAAGGTTTGAAGTAGCGTTGCTCGCC
 TTGAGCACGAAGAAGAAGCTCACTTTAGGAATCACTCTCTCTGCCTTCGCTCTTCTCTTA
 ACAGAGCTCGTGGCGGCGCGTGTTCACGCGCTCTAATAACACCGACAAAGACAAAAAC
 GCGATTGCCCGCGAGAAAAATCGAACTTTTGAAGTAACTCGAGTTCCCAAAGCGATTCCA
 TGTCTGAGGAAACAGAGCATGTAGTATCTGAAACAGAGGTTTGAAGTTGAAAGGTTTA
 ACGATACGTGATCTGTGTCAAAGGACGAGAAATCAACAAGTAAAGTTGGAGACTAAAA
 TCGAAGATTGTGAAGAAGTTGAGGAGTTACAATAAGAAGGATAAGAAGACGATGAAGATC
 AAAGAAGAGTCTTTGATTGAAGTCTCGAGTTTGGTTTTAGAAGATAAACCAAAGAAAATT
 GAGTCTGAGAGAGACGAAGAAGAAACGTTGAATCCTCCAGTGGTTGGATCAAACTGAAT
 GGGATTGTTCTGATCGTGATTGTGCTAACCGGTTTGTATGTGGGAAGGTCTTAGCTATT
 GTTCTGACACTATCATGTTTGGTTCTTAGATTAGGAGCAGTCAAAAAAGTTAATCTTTGC
 ATATAATTTTTTTTGTATTTTTTAACATGCTTGCATGTGAACTGTAAATTTTTCTCATT
 CATATGAAGGAGATTGGATTGAATGTTGAATACTAAA

>G1447 Amino Acid Sequence (domain in AA coordinates: 3-54, 124-156)

MSSLPWKKPKSSRILRFISEFQQSPFVETGFPTSLIDLFFKNRDLKKSPPSKRFQRIERO
 IRTAPNASSLSNQDTIFEKPSRIKTVRSKVEKVNCKVKSAAALKKNAIKNSVFSGSSEVV
 LMAFKVLIVALLALSTKKKLTGLITLSAFALLLTSLVAARVFTSRNNTDKDKNAIAREKI
 ETFDETRVPAIPCEPTEHVSETEVSKLGLTIRDLLSKDEKSTSKSWRLKSKIVKKL
 RSYNKKDKMTMKIEESLIEVSSLVLEDKPKKIESERDEEETLNPVVGSNLNGIVLIVI
 VLTGLLCGKVLAIVLTLSCVLRLGAVKKVNLCI*

>G323 (77..826)

CTGCTCATATCAGCCATTGACACAGTTGCTTTGGGTTTCCCTCAAACGGCGCCGATTGTC
 TGGATTTTGACCACTGATGGCCTTAGATCAATCTTTTGAAGATGCTGCTTTACTTTGGAGA
 ACTCTATGGAGAAGGTGCATTTTGTTCAGAGCAAGAAACCTGAACCCATTACAGTCTC
 GGTTCCTTCTGATGATACTGATGATTGCAATTTTACTGCAATATTTGCTTAGACTCGGT
 GCAAGAACCTGTTGTGACTCTCTGTGGTCACCTCTTTTGTGCTGGCCTTGATTCACAAATG
 GCTTGATGTACAGAGCTTCTCAACAAGTGATGAATACCAAGACATAGACAGTGTCTCTGT
 TTGTAATCTAAAGTTTCTCATTCTACTTTGGTTCTTTGTATGGTAGAGGCCGTTGTAC
 TACTCAGGAGGAAGGTAAAAACAGTGTGCCTAAAAGACCCGTAGGACCGGTTTATCGGCT
 TGAAATGCCGAATTCACCTTATGCAAGTACTGATCTGCGGTTATCACAACGGGTTTCATT
 CAATAGCCACAGGAAGGTTACTACCTGTCTCAGGGGTGATGAGCTCGAACAGTTTATC
 ATACTCTGCTGTTTTGGATCCGGTGATGGTGATGGTTGGAGAAATGGTAGCTACGAGGTT
 GTTTGGAACACGAGTGATGGATAGATTTGCGTATCCGGACACTTACAATCTCGCAGGGAC
 TAGCGGGCCGAGGATGAGAAGGCGGATAATGCAGGCAGATAAATCGCTGGGAAGAATCTT
 CTTCTTCTTTATGTGTTGTGTTGTTCTGTGTTCTCTTCTTGTGTTTAGGTTTTCATAGCTAG
 CTTGGTTCTGCTACTGTTCACTTCTTCTCAGG

>G323 Amino Acid Sequence (conserved domain in AA coordinates: 48-96)

MALDQSFEDAALLGELYGEGAFCKSKKPEPITVSPSDDTDDSNFDCNICLDSVQEPVV
 TLGHLFCWPCIHKWLVDVQSFSTSEYQRHRQCPVCKSKVSHSTLVPLYGRGRCTTQEEG
 KNSVPKRPVGPVYRLEMPNSPYASTDLRLSQRVHFNSPQEGYYPVSGVMSNSLSYSAVL
 DPVMVMVGEMVATRLFGTRVMDRFAYPDYTNLAGTSGPRMRRRIMQADKSLGRIFFFMFC
 CVVLCLLLF*

>G176 (41..1606)

AGAAGAAGAAGAAGAGTACCTCATACGTAAACCATTGATGGGCTCTTTTGATCGCCA
AAGAGCTGTTCCGAAATTCAAACAGCAACACCGTCACCGCTCCCTCTTTCTCCTTCGCC
TTACTTCACTATGCCTCCTGGCCTTACTCCCGCCGACTTTCTCGACTCTCCTCTTCTCTT
CACTTCCCTCAACATTTTGCCGTCTCCTACGACAGGCACATTTCCAGCGCAATCTCTGAA
CTATAACAATAACGGTTTGCTCATTGACAAAAATGAAATCAAATATGAAGACACAACCTCC
TCCCTTGTTCTTACCATCTATGGTAATCGAGCCTTTACCTCAACTGGATTATTCAAATC
CGAAATCATGTGCGAGTAACAAAACCTCTGATGACGGCTACAATTGGCGCAAAATACGGGCA
GAAGCAAGTCAAAGGAAGCGAAAACCCGAGGAGTTACTTCAAATGCACGTATCCAAATTG
TCTCACAAGAAGAAAGTAGAGACGTCTCTTGTGAAGGGTCAGATGATTGAGATTGTCTA
TAAAGGAAGCCACAATCATCCCAAGCCCCAATCCACGAAGCGATCATCTTCCACCGCTAT
AGCAGCACATCAGAACAGCAGTAATGGAGACGGTAAAGACATTGGTGAAGATGAAACAGA
GGCCAAGAGATGGAAGAGAGAAGAGAAATGTGAAGGAGCCAAGAGTGGTGGTTCAGACAAC
AAGTGATATAGACATTCTTGACGATGGCTACAGATGGAGAAAGTATGGTCAGAAAGTCGT
CAAGGGTAATCCAAATCCAAGGAGCTATTACAAGTGCACATTTACAGGATGTTTTGTAAG
GAAACACGTTGAAAGAGCATTTCAGATCCCAAGTCAGTGATCACAACCTACGAAGGAAA
ACACAAACACCAAATCCCGACCCCAAGAGAGGTCCAGTTTAAAGATCTGCTGCAATGGC
TTCTCCTCTTCTCCCAACTTCGACTACTCCTGATCAACTTCCCGCGGGCGATCCACAGTT
GCTGAGCTCTCTACGCGTCTCTTGTCCCGGTTCTAGCCACCGTCCGTACGCTTCTGC
AGATGCCAGACCTTGGGCGAGCTCGTTGACCGGTACGCTTTTCCCGGCCACCATCGCT
CTCGGAGGCAACGTACAGTAAGGAAGAACTTTTCTTATTTCCGAGCCAATTACATAAC
CTTAGTGGCAATCTTACTCGCCGCTCTCTGCTCAGCACCCCTTTTCGCTCTCTTCTCCT
CGCATCGCTGGCCGCTTCTTGCTTTTCTCTACTTTTCCGTCCGGCGGATCAGCCGTT
GGTCATTGGAGGACGACGTTCTCCGATCTTGAGACGCTAGGGATACTCTGCCTGTCCAC
TGTGGTGGTGGATGTTTATGACAGCGTTGGATCGCTCTTGATGTCCACTCTAGCAGTTGG
GATCATGGGCGTGGCCATCCACGGAGCGTTTCTGCTCCCGAAGACCTGTTCTTGAAGA
ACAAGAAGCCATTGGATCTGGACTTTTTCGATTCTTCAACAACAATGCCTCTAATGCAGC
TGCCGCTGCCATAGCCACCTCAGCAATGTACGCGTTCGAGTCTGAGATTGTTGAAGAGA
CTACATTCTTACACCGCATTTCCAAAGTGTGATATTTATTCATATTGAATTGTT
>G176 Amino Acid Sequence (domain in AA coordinates: 117-173,234-290)
MGSFDRQRAVPKFKTATPSPLPLSPSPYFTMPPLTPADFLDSPLLFTSSNILPSPTTGT
FPAQSLNYYNNGLLIDKNEIKYEDTTPPLFLPSMVTQPLPQLDLFKSEIMSSNKTSDDG
NWKYQKQKQVKGSENPRSYFKCTYPNCLTKKKVETSLVKQMIEIVYKGSNHPKQSTK
RSSSTAIAAHQNSSNGDGKDIGEDETAKRWKREENVKEPRVVVQTTSDIDILDDGYRWR
KYGQKVVKGPNPNRPSYKCTFTGCFVRKHVERAFQDPKSVITTYEGKHKHQIPTPRRGPV
LRSAMASPLPSTSTPDQLPGGDPQLLSLRLVLSRVLATVRHASADARPWAELVDRSA
FSRPPSLSEATSRRVKNFSYFRANYITLVAILLAASLLTHPFALFLLASLAASWFLYFF
RPADQLVIGGRTFSDLETLGILCLSTVVMFMTSVGSLLMSTLAVGIMGVAIHGAFRAP
EDLFLEEQAIGSGLFAFFNNNASNAAAAAIATSAMSRVRV*
>G174 (194..1585)
CCCAATTGAGATTGTTTCGATTTTCGATCTACGAGATTCTTACAAGAACATAAGCAGCTTC
GGTTTTTTGGGATTATCTTATTTGGTCGGATGATGATCTTCTCGATGTCTGTGCTAGGCT
TTGGGAATTAGATATATTTGGGGTTAAGCTCGAGTCTCTCCGGTTTTGAGTTTACTTGAG
TTTGTTAGTATTTATGGCTGAGGTGGGAAAAGTTCTGGCTAGTGATATGGAGTTAGACCA
TTCAAATGAGACTAAAGCAGTGGATGATGTTGTTGCCACTACTGATAAAGCGGAGGTCAT
ACCACTGGCTGTAAGTAACTAGAAACCGTTGTTGAAAGTTTGGAACTACTGACTGTAA
GGAGCTTGAAAACTTGTTCACATACGGTAGCTTCGCAGTCGGAAGTAGATGTTGCTTC
CCCGGTATCCGAGAAAGCACCGAAGGTTTCTGAAAGTAGCGGTGCATTATCTTTGCAGTC
TGGTTCGGAAGGGAATAGTCTTTTATTCTGTGAGAAGGTTATGGAAGACGGATACAACCTG
GCGGAAATATGGACAGAACTTGTGAAAGGAAATGAGTTTGTAAAGGAGCTATTACAGGTG
CACTCACCCCTAAGTGCAGAAAGCGAAACAGTTGGAACGGTCTGCGGGTGGACAAGTCGT
GGATACCGTTTACTTTGGGGAAACATGATCACCCAAAGCCTCTTGCTGGTGCTGTTCTCTAT
CAATCAGGATAAGCGAAGTGTGTTTACAGCTGTTAGTAAAGAGAAAAACATCTGGATC
CAGTGTTCAGACACTTCGTCAAACCGAACCACCAAGATCCATGGAGGATTACATGTTTC
AGTTATTCCACCAGCTGATGATGTGAAAACCTGATATTTACAATCAAGTAGGATAACGGG
GGACAACACTCACAAGGATTATAATAGTCCTACCGCAAAGCGAAGGAAGAAAGGAGGGAA
CATTGAGCTGAGTCCAGTGGAGAGGTCAACCAATGATTCACGCATTGTGGTTTCACTCA
GACTCTGTTTGATATTGTGAATGATGGTACCGATGGCGTAAATATGGTCAGAAATCAGT

AAAAGGCAGCCCATATCCAAGGAGCTACTATAGATGTTCAAGCCCTGGATGCCCCGTCAA
GAAACACGTAGAGAGGTCATCTCATGACACAAAGTTGCTTATAACAACCTACGAGGGAAA
ACACGACCACGATATGCCTCCAGGAAGAGTTGTTACTCATAATAACATGCTGGACTCGGA
AGTTGATGATAAAGAAGGAGATGCCAACAAGACTCCACAGAGCTCAACTCTTCAATCCAT
TACAAAAGACCAGCATGTCGAAGATCACTTAAGAAAAGAAAACGAAGACTAATGGCTTTGA
GAAAAGTCTTGATCAAGGTCCAGTTTGGATGAGAAGCTGAAGGAGGAAATAAAAGAGAG
ATCAGATGCAAACAAGATCACGAGCCAATCACGCCAAGCCGGAAGCAAAGTCAGATGA
TAAAACCACTGTTTGTCAAGAGAAGGCAGTAGGAACCTGGAGAGCGAGGAACAAAACC
CAAGACAGAGCCTGCCCAAAGCTAAGCATTCACTGTTGTACCGAGTGGTAATTTATATGG
CTGTTTTAACATAGATTAGTACAGGCGATATGGTTATAGACTGTACAGTTGTTGTTTCAGG
CGGGACCAGATTTAGATTAGTGTTTAATGGAATAGTATGCTTTAATACCTTTATGTAACC
ACTTCCATTTGGTTCAAATAAGAGTTACAGGAAGAGAAGGTAACACAACAAGAGCCCTTC
TTGTTGATGGAGCCTGTGTAATAGTTGTAGCATGGGGATGTATATGATTGATTCAACC
TTATTAATGGTTATGAGACAAAACCTATC

>G174 Amino Acid Sequence (domain in AA coordinates: TBD)

MAEVGKVLASDMELDHSNETKAVDDVVATTDKAEVIPVAVTRTETVVESLESTDCKELEK
LVPHTVASQSEVDVASPVSEKAPKVSESSGALSLSQSGSEGNPFIREKVMEDGYNWRKYG
QKLVKGFNFVRSYRCHPNCKAKKQLERSAGGQVVDTVYFGEHDHPKPLAGAVPINQDK
RSDVFTAVSKEKTSGSSVQTLRQTEPPKIHGGLHVSVIPADDVKTDISQSSRITGDNTH
KDYNSTAKRRKKGGNIELSPVERSTNDSRIVVHTQTLFDIVNDGYRWRKYGQKSVKGS
YPRSYRCSSPGCPVKHVERSSHDTKLLITTYEGKHDHDMPPGRVVTHNMLDSEVDDK
EGDANKTPQSSTLQSIKTDQHVEDHLRKKTKTNGFEKSLDQGPVLDEKLKEIKERSDAN
KDHAANHAKPEAKSDDKTTVCQEKAVGTLESEEQPKTEPAQS*

>G715 (1..705)

ATGGATACCAACAACAGCAACCACCTCCCTCCGCCGCCGGAATCCCTCCCTCCACCACCT
GGAACCACCATCTCCGCCGCAGGAGGAGGAGCTTCTTACCACCACCTTCTCCAACAACAA
CAACAACAGCTCCAACCTATTCTGGACCTACCAACGCCAAGAGATCGAACAAGTTAACGAT
TTCAAAAACCATCAGCTTCCACTAGCTAGGATAAAAAAGATCATGAAAGCCGATGAAGAT
GTTTCGTATGATCTCCGCAGAAGCACCAGATTCTCTTCGCGAAAAGCTTGTGAGCTTTTCATT
CTCGAGCTCAGCATCAGATCTTGGCTTCACGCTGAGGAGAATAAACGTCGTACGCTTCAG
AAAAACGATATCGCTGCTGCGATTACTAGGACTGATATCTTCGATTTCTTGTGATATT
GTTCTAGAGATGAGATTAAGGACGAAGCCGCAGTCCTCGGTGTTGGAATGGTGGTGGCT
CCTACCGCGAGCGCGCTGCCTTACTATTATCCGCCGATGGGACAACCAGCTGGTCTCTGGA
GGGATGATGATTGGGAGAGACCAGCTATGGATCCGAATGGTGTGTTATGTCCAGCCTCCGTCT
CAGGCGTGGCAGAGTGTGTTGGCAGACTTCGACGGGGACGGGAGATGATGCTCTTATGGT
AGTGGTGAAGTTCCGGTCAAGGGAATCTCGACGGCCAAGGGTAA

>G715 Amino Acid Sequence (domain in AA coordinates: 60-132)

MDTNNQQPPPSAAGIPPPPPGTTTISAAGGASVYHLLQQQQQLQLFWTYQRQIEQVND
FKNHQLPLARIKKIMKADEDVRMISAEAPILFAKACELFILELTIRSWLHABENKRRTLQ
KNDIAAAITRTDIFDPLVDIVPRDEIKDEAAVLGGGMVVAPTASGVPIYYPMPGQAPGPG
GMMIGRPAMPDPNGVYVQPPSQAWQSVWQTSSTGTGDDVSYSGSGSGQNLGQGG*

>G588 (196..1599)

ATCTGAAGTGAACCAAGCTCAGGTTTTGCTCTCTCTTTGATCATTCCTTTCTCAGCAATA
TAAATTAGAGTTATATCCTTTATAAAGGATTTTGCTTTTTCACCAACAAACCCTAAATTC
GGTGTCTCAGCAAGAATCACGTGATTCTCGTTCCTCTTCCTCAGCAAACCCATCATCTTC
TATCTCATTTGAGAAATGGGTCAAAAGTTTTGGGAGAATCAAGAAGATCGAGCGATGGTT
GAATCCACCATAGGCTCTGAAGCTTGGCACTTTTTCATCTCAACAGCTTCAGCTTCCAAC
ACTGCCTTGTCCAAGCTTGTCTCACCACCAAGTGATTCCAATCTCCAACAAGGGTTACGT
CACGTTGTTGAAGGATCTGATTGGGATTATGCTCTTTTCTGGCTAGCGTCCAACGTTAAT
AGCTCTGATGGTTGTGCTCTTGATCTGGGGAGATGGTCATTGCCGTGTCAAAAAGGGTGCT
TCAGGTGAGGATTACTCTCAGCAAGATGAGATCAAAAGACGTGTGCTTCGCAAGCTTCAC
TTGTCTGTTCTGTTGGTTGATGAAGATCATCGTTTGGTGAAATCAGGAGCTCTTACTGAT
CTCGACATGTTTTATCTGGCTTCTTTGTACTTTTCTTTAGGTGTGATACCAATAAGTAC
GGTCTGCTGGAACCTATGTGTCTGGGAAGCCTCTTTGGGCTGCAGATTTGCCTAGCTGC
TTGAGTTATTATAGGGTTAGGTCTTTCTTAGCTAGGTGAGCTGGTTTTTCAGACTGTGTG
TCTGTACCAGTGAATTCTGGAGTTGTGGAGCTTGGTTCTTTAAGACATATTCCAGAAGAT
AAGAGTGTGATTGAGATGGTGAAATCAGTGTGTTGGTGGGTCTGACTTTGTTTCAGGCTAAA

GAAGCTCCTAAATCTTTGGTCGACAGCTGAGTCTTGGTGGAGCAAAACCTCGGTCTATG
 AGTATTAATTTCTCCCGAAGACCGAGGATGACACGGGTTTCTCATTGGAATCGTATGAG
 GTGCAAGCGATCGGAGGCTCTAATCAAGTGTATGGTTATGAGCAAGGGAAAGATGAGACA
 TTGTATCTAACTGACGAGCAAAAGCCGAGGAAGAGAGGGAGAAAACCGACAAATGGAAGA
 GAAGAGGCTCTAAACCATGTGGAAGCGGAACGGCAGAGGAGGGAGAACTGAACAGAGA
 TTCTACGCTTTTGAGAGCGGTGGTGCCCTAACATCTCCAAGATGGACAAGGCTTCGCTCCTT
 GCAGACGCAATCACTTACATCACGGATATGCAGAAGAAAATCAGGGTGTATGAAACAGAG
 AAGCAGATAATGAAGAGGAGGGAGAGTAATCAGATAACTCCAGCAGAGGTTGATTATCAA
 CAGAGGCATGATGATGCAGTTGTAAGGCTAAGCTGTCCGTTGGAACTCATCCAGTTTCA
 AAGGTGATACAAACGTTGAGGGAGAATGAAGTTATGCCTCATGATTCCAACGTGGCCATC
 ACAGAGGAGGGTGTGGTTACACATTCACTCTCCGGCCTCAGGGTGGCTGCACCGCTGAG
 CAGTTGAAGGACAAGCTCCTTGCCTCTCTATCACAGTAACATATCACAGCAGTAACGTCTA
 TGTAATAAGTGTAAACCGTGTGGAGGTTGTATCAATGTACTATTGCAAGCCAACCAAAAA
 AAACCTCCAGCTTAGTAGGATCGTGTAAATTTCTTATATGTAATGTTGAGATTGTCTTT
 TACATATAAAGATTGA

>G588 Amino Acid Sequence (domain in AA coordinates: 309-376)
 MGQKFVENQEDRAMVESTIGSEACDFFISTASASNTALSKLVSPPSDNLQQLRHVVEG
 SDWDYALFWLASNVNSDGCVLWGDGHCVRVKGASGEDYSQQDEIKRRVLRKLHLSFVG
 SDEHDRLVKSGALTDLDMFYLASLYFSFRCDTNKYGPAGTYVSGKPLWAADLPSCLSYYR
 VRSFLARSAGFQTVLSVPVNSGVVELGSLRHIPEDKSVIEMVKS VFGGSDFVQAKEAPKI
 FGRQLSLGGAKPRSMSINFSPKTEDDTGFSLESYEVQAIIGSNQVGYEQKDETLYLTD
 EQKPRKRGRKPANGREEALNHVEAERQRREKLNQRFYALRAVVPNISKMDKASLLADAIT
 YITDMQKKIRVYETEKQIMKRRESNQITPAEVDYQQRHDDAVVRLSCPLETHPVSKVIQT
 LRENEVMPHDSNVAITEEGVHTFTLRPQGGCTAEQLKDKLLASLSQ*

>G1758 (69..677)
 GTCCCTCCTCTTAGCTTCAACCGCCGGAATAAACAACCTTCTTGAAAAAAGAGA
 AACTAAAAATGAATATCCTTCAAACCTAACCCTAGCTCCACAGATTTCACTGAATTTT
 TCAAGTTCGATGATTTTGACGATACTTTGGAGAAGATCATGGAAGAAATCGGCCGTGAGG
 ACCACTCGTCGTCACCGACTTTGAGTTGGAGTTCATCGGAAAAGTTAGTGGCTGCAGAAA
 TCACAAGCCCGCTTCAAACAAGCCTAGCTACCTCACCTATGAGCTTTGAAATAGGTGACA
 AAGATGAAATCAAAAAGAGGAAGAGACACAAAGAAGATCCGATTATTCACGTCTTCAAAA
 CGAAATCATCAATTGATGAAAAGGTTGCTTTAGATGATGGGTATAAATGGAGGAAATACG
 GAAAGAAGCCGATAACGGGTAGTCCATTTCCAAGGCATTATCACAGTGTTTCGAGCCAG
 ATTGCAACGTGAAGAAGAAGATCGAAAGAGATACGAACAATCCAGATTACATATTGACAA
 CATACGAAGGTAGACATAACCACCCAAGCCCTTCTGTAGTTTATTGTGATTACAGACGACT
 TTGATCTTAACCTCTCTCAACAATTGGTCCTTTCAGACGGCAAATACGTATAGTTTCTCTC
 ATTCTGTCCATATTGATCGATCGTAGTTACAAGTTTGTGTATATAGATGTATATATATA
 TATACCAATTACCATCGTAATCACGTCTCACATGTAACGTACATATATCTTGTTT
 GGGGTTTCGTTTGTAAATGTATTGAATTGGTGGAGGTAGAAATGGAAGTCATCTTGATAGT
 TGTACTTGTATGTAAGTTTGATAGTCATTTTTTATAAAGTAACTAATTTGTACAA

>G1758 Amino Acid Sequence (domain in AA coordinates: TBD)
 MNYPSPNPSPSTDFTEFFKFDDFDDTFEKIMEEIGREDHSSSPTLSWSSSEKLVAEITS
 PLQTSLATSPMSFEIGDKDEIKRKRHKEDPIIHVFKTKSSIDEKVALDDGYKWRKYGKK
 PITGSPFPRHYHKSSPDCNVKKIERDTNPNPDYILTTYEGRHNHPSVSVYCDSDDFDL
 NSLNNWSFQTANTYSFHSAPY*

>G2148 (66..737)
 GTCTCTAATATAAGCTTGAACGTTGCTATATATAAATGTAAAGGCGAACGCATAAGAAAA
 GAAAAATGGAGAATGAAGCTTTTGTAGATGGTGAATTGGAGTCTCTTTTGGGGATGTTCA
 ACTTTGATCAATGTTTCATCTAACGAATCGAGCTTTTGAATGCTCCAAATGAGACTGATG
 TTTTCTCTTCTGATGATTTCTCCATTTGGTACAATCTGCAAAGTAACTATGCGGCCG
 TTCTTGATGGTTCCAACCACCAAACGAACCGAAATGCGACTCAAGACAAGATCTGTTGA
 AACCAAGGAAGAAGCAAAAGTTAAGCTCGGAAAGCAATTTGGTTACCGAGCCTAAGACTG
 CTTGGAGAGATGGTCAAAGCCTAAGCAGTTATAATAGTTTCAAGATGATGAAAAGGCTTTAG
 GTTAGTGTCTAATACATCAAAAAGCCTAAAACGCAAAGCGAAAGCCAACAGAGGGATAG
 CTTCGATCCTCAGAGCCTATACGCTAGGAAACGAAGAGAAAGGATAAACGATAGGCTAA
 AGACATTGCAGAGCCTAGTTCCCTAATGGGACAAAGGTGATATAAGCACAAATGCTGGAAG
 ATGCTGTCCATTACGTGAAGTTCCTGCAGCTTCAATCAAGCTCTTGAGTTCAGAAGATC

TATGGATGTATGCACCTCTTGCTCACAATGGTCTGAATATGGGACTACATCACAATCTTT
TGTCTCGGCTTATTTAAGACAAAATCATTGGAATAACATAACTTACAGTACTTGTTTTTT
TTCTCGTTCTATATTCATGATTATGGTTATTTTTTGTGTTGAGTTGTTCAATTTTTCTGTC
TATTGCGTTCTATGAACCTGACACTCTTTTTGTAATTATTATATGCTAAAGACAATTGG
ACTAACAGCATTTTAATAAAAAAAAAAAAA

>G2148 Amino Acid Sequence (conserved domain in AA coordinates:130-268)

MENEAFVDGELESLLGMFNFDQCSSNESSFCNAPNETDVFSSDDFFPFGTILQSNYAAVL
DGSNHQTNRNVDNRQDLLKPRKKQLSSESNLVTEPKTAWRDGQSLSSYNSSDDEKALGL
VSNTSKSLKRKAKANRGIASDPQSLYARKRRERINDRLKTLQSLVPNGTKVDISTMLEDA
VHYVKFLQLQIKLLSSEDLWMYAPLAHNLNMGHLHNLRLI*

>G2379 (52..798)

CGCCGTCACCTCTCTCCCGGTGCCGCACATTAGCAACACTACTCCCGACGAATGGAGACG
ACGACGCCGCGAGTCAAAATCAAGTGTGTCCCACCGACCGCGTTGGGAAGAGAAGACTGG
TGGAGTGAGGAAGCGACGCGACGCTGGTAGAAGCCTGGGGCAATCGTTACGTCAAGCTG
AACCACGGAAATCTCCGGCAGAAATGACTGGAAAGACGTCGCCGACGCCGTTAACTCTAGA
CACGGTGATAACAGCCGTAAGAAGACCGACTTACAGTGTAAAGAACGGGTCGATACTTTG
AAGAAGAAGTACAAAACAGAGAAAGCTAAACTCTCGCCGTCGACTTGGCGTTTCTATAAC
CGCCTCGATGTTCTAATCGGTCCCGTTGTGAAGAAATCGGCTGGCGGAGTTGTCAAATCA
GCGCCTTTTAAAGATCATCTGAATCCAATCGGATCGAATCTACTGGAAGCTCTCTTGAA
GATGATGATGAGGATGATGATGAGGTTGGTGATTGGGAATTCGTTGCTAGGAAGCATCCT
CGTGTGGAAGAGGTAGATCTGAGTGAAGGATCAACGTGTAGGGAAGTAGCTACGGCGATT
CTCAAGTTTGGAGAAGTTTACGAGAGAATTGAAGGAAGAAGCAACAGATGATGATTGAG
TTGGAGAAGCAGAGAAATGGAAGTGACAAAGGAGGTAGAGTTAAACGAATGAACATGTTG
ATGGAGATGCAGTTAGAGATTGAGAAATCAAAGCACCGGAAACGCGCAAGTGCTTCAGGT
AAGAAGAACTCACATTAGG

>G2379 Amino Acid Sequence (domain in AA coordinates:19-110, 173-232)

METTTTPQSKSSVSHRPLGREDWWSEATATLVEAWGNRYVKLNHGNLRQNDWKDVADAV
NSRHGDNRSRKTDLQCKNRVDTLKKKYKTEKAKLSPSTWRFYNRLDVLIGPVVKKSAGGV
VKSAPFKNHLNPTGSNSTGSSLEDDDDDDDEVGDWEFVARKHPRVEEVDLSEGSTCRELA
TAILKFGEVYERIEGKKQMMIELEKQRMVTEKEVELKRMNMLMEMQLEIEKSKHRKRAS
ASGKKNSH*

>G1462 (63..1031)

CGTCGACCATTTCTTGCGATTGATCTTTCTCTAGATAATTTTTTTTGATCGATTTAGTTTCA
TTATGGAGGACGACGACGAGCTTATGATCTAATCAAACACGAAGTGTATCTCAGAAG
ACGAAGTAATAATCTCAGTTATCTGAAGGGTATGGTCGTTAACGGAGATTCTTGCCAG
ATCACTTCATCGAAGACGCAACGTTTACCAAGAATCCAGATAAGGTGTTCAATTCTG
AGAGACCTAGATTCTGTGATCGTTAAACCACGAACAGAGGCTTGTGGTAAACCGATGGAT
GTGATTCCGGGTTGCTGGAGGATCATTGGTCGTGATAAACTGATAAAGTCGGAGGAGACTG
GGAAGATTCTAGGGTTCAAGAAGATACTCAAGTTTGCCTAAAGAGGAAACCTATAGACT
ACAAGAGAAGTTGGGTAATGGAAGAGTATAGGCTTACCAATAACTTGAAGTGAAGCAAG
ATCATGTGATTTCGAAAATTCGGTTTATGTTTGAAGCTGAAATTAGTTTCTTGCTAAGCA
AGCATTTCTACACTACATCAGAATCGGTTCTTGAATGAGCTGTTGCCATCTTATGGAT
ATTATTTATCCAATACACAAGAGGAGGATGAATTTTATCTGGACGCGATAATGACTTCGG
AAGGAAACGAGTGGCTAGCTACGTTACCAACAACGTGTACTGTCTGCATCCATTGGAGC
TTGTGGATCTTCAAGATCGGATGTTTAAATGATTACGGAACCTGCATCTTCGCTAACAGA
CTTGTGGTGAACTGATAAATGCGATGGTGGTTACTGGAAGATCCTGCACGGTGATAAGC
TGATCAAGTCAAATTTCCGAAAGGTCATTGGTTTCAAGAAGGTATTTGAGTTCTATGAAA
CGGTGAGACAAATATATCTTTGTGATGGAGAAGAAGTGACGGTAACCTTGGACTATACAAG
AGTATAGGCTTAGCAAAAACGTAAGCAGAATAAAGTGTTGTGCGTTATCAAGTTGACTT
ATGATAGATAGGATACTTTACTTTGGTTTGTGATCATCTTAGTATCTTACGAATATTC
TAGATACACACATCTATAGGCGACCGCTCTAGACAGGCTCGTACCG

>G1462 Amino Acid Sequence (domain in AA coordinates: TBD)

MEDDDAAYDLIKHELLYSEDEVIIISRYLKMVVGNDSPDHFIEDANVFTKNPKVFNSE
RPRFVIVKPRTEACGKTGDCDSGCWRIIGRDKLIKSEETGKILGFKILKFCLKRKPIDY
KRSWVMEEYRLTNLNLWKQDHVICKIRFMFEAEISFLLSKHFYTTSESVLENELLPSYGY
YLSNTQEEDEFYLDAIMTSEGNEWPSYVTNNVYCLHPLLELVDLQDRMFNDYGTICIFANKT
CGETDKCDGGYWKILHGDKLIKSNFGKVIKVFYETVRQIYLCDGEEVTVTWTIQE

YRLSKNVKQNKVLCVIKLTDR*

>G1211 (44..1120)

TGAAACCTAGATTTCTGCAACTGAATTCCTAATTCGAAAAAGAATGGAGGGTTCGTCGTC
GACGATAGCAAGGAAGACATGGGAAC TAGAGAACAGCATTCTAACAGTAGACTCACCTGA
TTCAACCTCCGACAACATCTTCTACTACGACGATACTTACAGACTAGGTTCCAGCAAGA
GAAACCGTGGGAGAATGATCCTCACTACTTTAAACGAGTCAAGATCTCAGCGCTCGCTCT
TCTTAAGATGGTGGTTCACGCTCGCTCTGGTGGTACAATTGAAATAATGGGTCTTATGCA
AGGTAAGACCGATGGTGATACTATCATTTGTTATGGATGCTTTTGCTTTACCAGTGGAAGG
TACTGAGACAAGGGTTAATGCTCAGGATGATGCTTATGAGTACATGGTTGAGTATTACACA
GACCAACAAGCTCGCGGGGCGGCTGGAGAATGTTGTTGGATGGTATCACTCTCACCCTGG
ATATGGATGCTGGCTCTCCGGTATTGATGTTTCTACGACGACGCTTAACCAACAGCATCA
GGAGCCATTTT TAGCTGTTGTTATTGATCCCAAGGACTGTTTCAGCTGGTAAGGTTGA
GATTGGTGTCTTTCAGAACATACTCTAAAGGATATAAGCCTCCAGATGAACCTGTTTCTGA
GTATCAAAC TATTCCTTTAAATAAGATTGAGGACTTTGGTGTTCAGTGCAGAACAGTACTA
TTCATTAGATGTCATTATTTCAAGTCATCTCTTGATTCTCACCTTCTGGATCTACTATG
GAACAAGTACTGGGTGAACACTCTTTCTCTCTCCACTGCTGGGTAAATGGAGACTATGT
TGCTGGACAAATATCAGACTTAGCTGAGAAGCTTGAGCAAGCCGAGAGTCATCTGGTTCA
GTCTCGCTTTGGAGGAGTTGTGCCATCATCCCTTCATAAGAAAAAGAAGATGAGTCTCA
ACTAATAAGATAACTCGGGATAGCGCAAAGATAACTGTGGAACAGGTCCATGGACTAAT
GTGCGAGGTCAATAAAGATGAATTATTCAACTCAATGCGTCAGTCCAAACAACAAATCTCC
CACTGACTCGTCGGATCCAGACCCTATGATTACATATTGAAGTTGCTCTCTTTTGGTTT
CTANTTTTGGATTGACCCATCATTTGTTGCTCTTTCATTTATTTTCTGTTGTGTAAAGAA
TTATAATGNCNGCGCAATTCGCGGCCGCTAAAAAANACAGGAAATTGAAAANAATTCTN
NCCATTCCAACATCTTTATTTAATATTATCTCTCNATTATATAATATTCAAACATCCCT
ANTANCTTCATTTGACCGTCCCCCTCCCTCCCGTGTTCNTTGGTGTCTGGCCCC

>G1211 Amino Acid Sequence (domain in AA coordinates: 123-179)

MEGSSSTIARKTWELENSILTVDSPDSTSDNIFYDDTSQTRFQQEKPWENDPHYFKRVK
ISALALLKMOVHARSGGTIEIMGLMQKTDGDTIIVMDAFALPVEGTETRVNAQDDAYEY
MVEYSQTNKLAGRLNENVGWYHSHPGYGCWLSGIDVSTQTLNQHQEPFLAVVIDPRTV
SAGKVEIGAFRTYSKGYKPPDEPVSEYQTIPLNKIEDFGVHCKQYYSLDVITYFKSSLDSH
LLDLLWNKYWVNTLSSSPLLGNGDYVAGQISDLAEKLEQAESHLVQSRFGGVVPSSLHKK
KEDESQTLKITRDSAKITVEQVHGLMSQVIKDELFNSMRQSNKSPDSSDPDPMITY*

>G1048 (5..892)

GACCATGGCGGAGGAATTTGGAAGCATAGATTTACTCGGAGATGAAGATTTCTTCTTCGA
TTTCGATCCTTCAATCGTAATTGATTCCTTCCGGCGGAGGATTTTCTTCAGTCTTCACC
GGATTCTAGGATCGGAGAAATCGAGAATCAATTGATGAACGATGAGAATCATCAAGAGGA
GAGTTTTGTGGAATTGGATCAGCAATCGGTTT CAGATTT CATAGCGGATCTACTCGTTGA
TTATCCAAC TAGCGATTTCTGGCTCCGTTGATTGGCGGCTGATAAAGTTCTAACCGTCGA
TTCTCCCGCCGCCGCTGATGATTCCGGGAAGGAGAATTCCGATTGTTGTTGAGAAGAA
GTCTAATGATTCCTGGTAGCGAGATT CATGATGATGATGACGAAGAAGGAGACGATGATGC
TGTGGCTAAAAAACGAAGAAGGAGAGTAAGAAATAGAGATGCGGCGGTTAGATCGAGAGA
GAGGAAGAAGGAATATGTACAAGATTTAGAGAAGAAGAGTAAGTATCTCGAAAGAGAATG
CTTGAGACTAGGACGTATGCTTGAGTGCTTCGTTGCTGAAAACCACTCTACGTTACTG
TTTGCAAAAGGGTAATGGCAATAATACTACCATGATGTGCAAGCAGGAGTCTGCTGTGCT
CTTGTTGGAATCCCTGCTGTTGGGTTCCCTGCTTTGGCTTCTGGGAGTAAACTTCATTTG
CCTATTCCCTTATATGTCCACACAAAGTGTGCCTCTACGTCCAGAACCAGAAAAGCT
GGTTCTAAACGGGCTCGGGAGTAGTAGCAAACCGTCTTATACCGGCGTTAGTCGGAGATG
TAAGGGTTTCGAGGCCTAGGATGAAATACCAAATCTTAACCTTGCAGCGTGACAACGCCT
TTTTTAACTGCTTCTTTTGCGCATTTT GAGTTGTAGATGAGTGTCTTTTAGTTTTCTCTC
TCTTGTTTTGTATTTCGCTGTTGAAAGTTTCTGTCTAATATCGATAAGTTAACAGTGAA
AAAAAAAAAAAAAA

>G1048 Amino Acid Sequence (domain in AA coordinates 138-190)

MAEEFGSIDLLGDEDFFFDFDPSIVIDSLPAEDFLQSSPD SWIGEIEENQLMNDENHQEES
FVELDQSVSDFIADLLVDYPTSDSGSVDLAADKVLTVDSPAAADDSGKENS DLVVEKKS
NDSGSEIHDDDDDEEGDDDAVAKRRRRVRNRDAVRSRERKKEYVQDLEKKS KYLERECL
RLGRMLECFVAENQSLRYCLQKNGNNTTMSKQESAVLLLESLLLGSLLWLLGVNFICL
FPYMSHTKCCLLRPEPEKLVNLGLSSSKPSYTGVSRRCKGSRPRMKYQILTLAA*

>G986 (31..846)

CATTAAATTGGCTCCTGTGAACCTAAATTTATGGACTATGATCCCAACACCAATCCGTTT
GACCTTCATTTCTCCGGTAAACTTCCGAAAAGAGAAGTCTCGGCTTCAGCTTCTAAAGTT
GTAGAGAAGAAATGGTTAGTGAAAGATGAGAAGAGAAATATGCTACAAGATGAAATAAAC
CGGGTTAATTCGGAGAACAGAAGCTAACCGAAATGTTAGCAAGAGTCTGTGAGAAGTAC
TATGCTCTTAATAATCTTATGGAGGAGTTGCAGAGTCGAAAGAGTCTGAAAGTGTTAAC
TTTCAGAACAAACAGCTAACGGGGAAACGAAAACAAGAAGTCTGATGAGTTTGTTAGCTCC
CCAATTGGACTCAGTCTCGGACCAATCGAGAACATCACCAACGATAAAGCGACGGTTTCA
ACCGCTTACTTTGCTGCTGAGAAGTCTGACACAAGCTTGACTGTGAAAGATGGATATCAA
TGGAGGAAATACGGGCAAAAGATTACGAGAGATAATCCATCTCCTAGAGCTTACTTCAGA
TGCTCGTTTTACCGTCTTGCTAGTCAAGAAGAAGGTGCAACGAAGTGCAGAAGATCCA
TCTTTCTTGGTAGCCACTTACGAAGGGACACATAACCACACCGGACCACATGCAAGTGTG
TCCAGGACAGTGAAACTTGATCTAGTTCAAGGTGGGCTTGAACCAAGTTGAGGAAAAGAAA
GAGAGAGGGACGATTCAAGAGGTTTTGGTGCACAAATGGCTTCTTCGTTGACCAAAGAT
CCTAAGTTCACTGCAGCTCTTGCGACTGCTATTTCCGGGAGATTGATAGAGCATTCAAGA
ACATGAAAGTTCTCTAGAACATGTATATTTCTGTTTGTCTATTTTGTGCTCATTCCCT
AGTAAAAAGGTAAAGATTTGTTTGATCTTGATTAGGAGGCATAGATGTCAATTTTAATGT
GTGTGTATATAATTACATCAATCTAAGTATCCAAAAAGGGTCACCCCCATTTTATCTTA
TG

>G986 Amino Acid Sequence {domain in AA coordinates: 146-203}

MDYDPNTNPFDLHFSGLPKREVSAASASKVVEKKWLVKDEKRNMLQDEINRVNSENKKLT
EMLARVCEKYYALNNLMEELQSRKSPESVNFQNKQLTGKQKQELDEFVSSPIGLSLGPIE
NITNDKATVSTAYFAAEKSDTSLTVKDGQWRKYQKQITRDNPSPRAYFRCSFSPSCLVK
KKVQRSAEDPSFLVATYEGTHNHTGPHASVSRVTKLDLVQGGLEPVEKKERGTTIQEVLV
QQMASSLTKDPKFTAALATAISGRLEHSRT*

>G789 (259..1593)

GGCAAGAAGAACCTTAGCCTCTCTTTCTTCTTCTCTCTCTCTCTCTGTGGTACTGTT
CTGTTTCAACTTTACTCCCTCAGTTTCAGAACAATTCCTATCTAGAAGAGAGATAAAAC
CGAGAAGGTTTTGGAGATAGAATCTTTGTTCTTCTTTTGTCCCTCCTTGCTCGATTTTT
GTTACGTGTGAAGCAATAAAAAAACTGATATAGCTAAATCTTCCATCCATTAGAGGC
TTCTAAATCTGATCTGACATGGAACAAGTGTGCTGATTGGAATTTTGAAGATAATTTT
CACATGTCCACTAATAAAAGATCAATCAGACCAGAAGATGAATTAGTGGAGCTATTGTGG
AGAGATGGTCAAGTGGTTTTACAAAGCCAAGCTCGTAGAGAACCGTCAGTCCAAGTCCAA
ACCCACAAACAAGAAACCTTAAGAAAACCCCAACAATATTTTCTTGACAACCAAGAAACA
GTACAAAAGCCTAACTACGCTGCTCTAGATGATCAAGAAACCGTCTCCTGGATACAATAC
CCTCCGGATGACGTCATCGACCCCTTTCGAATCCGAGTCTCCTCTCATTTCTTCTCTCG
ATCGATCACCTCGGAGGTCTTGAGAAGCCACGAACGATCGAAGAGACAGTTAAGCATGAG
GCTCAAGCCATGGCTCCTCCTAAGTTTAGATCCTCGGTTATAACAGTCGGACCGAGTCAT
TGCGGCAGCAACCAGTCAACAAATATTCATCAGGCCACTACACTTCCGGTTTCTATGAGT
GATAGAAGCAAGAACGTGGAAGAAAGACTTGACACTTCGTCAGGTGGCTCCTCCGGTTGC
AGCTATGGAAGGAACAACAAGAAACCGTTAGTGAACAAGTGTAACCATTGACCGTAAA
AGAAAACATGTTATGGATGCTGATCAAGAATCTGTGCTCAATCAGATATAGGTTTGACC
TCAACCGATGATCAAAACCATGGGTAAACAATCGAGCCAACGGTCAGGATCTACTCGAAGA
AGCCGTGCAGCTGAAGTTTATAATCTCTCAGAAAGGAGGAGGAGAGATCGGATCAATGAA
AGAATGAAAGCTCTTCAAGAACTCATACCTCACTGCAGCAGAACAGATAAAGCTTCGATA
TTGGATGAAGCAATTGATTACTTAAATCACTTCAAATGCAACTCCAAGTGATGTGGATG
GGAAGTGAATGGCGGCGGCGGAGCAGCAGCAAGTCCGATGATGTTTCCCGGGGTA
CAATCATCTCCATACATTAATCAGATGGCTATGCAAAGTCAGATGCAATTGTCTCAATTC
CCGGTTATGAACCGGTCCGCTCCGCAGAACCATCCCGGTTTAGTATGTCAAAACCCGGTA
CAGTTGCAGCTCAAGCACAGAACCAATCTTATCGGAGCAGCTCGCTAGGTACATGGGC
GGGATTTCCCGAGATGCCGCGGCGGAAATCAGATGCAGACCGTGCAACAACAACAGCG
GACATGTTGGGATTTGGATCTCCGGCGGGACCGCAAAGTCAACTGTCCGCACCGGCGACC
ACCGACAGTCTTCATATGGGTAAATAGGCTGACTTGGCATATAGTTTTCTCCGAAATT
ATTCTTCTTACAGTTGGTGATTGTTATTTATTTTGGTCGCCTAAGCAAGCATAAAAGCT
AAGTCAAATGTATTATAGAGATCTAATAAGTTAGTCTCATACTTATAACTTATTTTTAAA
CAGTTGAATTATAGTATCAATCAAGTGTGGGAACCTAAGATCATACATGTGTCAATAC
TTTTATATTTGTTCTCAAGGTTTCATCAGAAAAACAAATAAAAAGGATAGACTAGGCCTG

CATTTGACATTATCATGGGCTTTTTTGGGTCTATGAATATGAACATTAACCCC
>G789 Amino Acid Sequence (domain in AA coordinates: 253-313)
MEQVFADWNFEDNFHMSTNKRIRPEDELVELLWRDQGVVLQSQARREPSVQVQTHKQET
LRKPNNIFLDNQETVQKPNYAALDDQETVSWIQYPPDDVIDPFSEFSSHFSSIDHLGG
PEKPRTIETVKHEAQAMAPPKFRSSVITVGP SHCGSNQSTNIHQATTLPVMSDRSKNV
EERLDTSSGGSSGCSYGRNNKETVSGTSVTIDRKRKHVMDADQESVSQSDIGLTSTDDQT
MGNKSSQSRSGSTRRSRAAEVHNLSERRRRDRINERMKALQELIPHCSRDKASILDEAID
YLKSLQMLQVMWMGSGMAAAAAAASPMMPFGVQSSPYINQMAMQSQMQLSQFPVMNRS
APQNHPLVLCQNPVQLQLQAQNLSEQLARYMGGIPQMPPAGNQMQTVQQPADMLGFG
SPAGPQSQLSAPATTDLSLHMKIG*

>G2085 (1..930)

ATGTTTGGTGCCTTCGATTATCCCAAATAACCAGATTGGTACCGCCTCTGCTTCCGCT
GGTGAAGACCATGTCTCTGCTCCGCTACGTCCTGGTCACATTCTTACGACGATATGGAA
GAAATCCCTCATCTGACTCTATCTATGGTGTCTGCTCCGATTGATTCCCGATGGCTCT
CAATGGTTGCTCACCGATCCGATGGCTCTGAATTACTTGTCTTCGCCACCGGAAGGG
GCGAATCAGCTTACGATCTCGTTCCGTGGACAAGTTTACGTTTTTGATGCCGTTGGTGTCT
GACAAGGTGGATGCTGTGTGTGCTGCTGTTGGGTGGTTCTACTGAGCTTGCTCCTGGTCCG
CAGGTGATGGAAGTACTCAACAGCAGAATCATATGCCTGTTGTAGAATATCAGAGCCGC
TGTAGCCTTCCGCAACGGGCACAATCCTTGGATAGGTTTCGGAAGAAGAGGAATGCTAGA
TGTTTCGAGAAGAAAGTAAGATACGGTGTTCGCCAAGAAGTTGCCTTAAGAATGGCACGT
AATAAAGGTCAATTCACCTCTTCAAAGATGACAGATGGGGCTTATACTCTGGCACAGAT
CAAGATTCTGCCAAGATGATGCCCATCCAGAAATATCGTGTACTCATTGCGGCATTAGT
TCCAAATGTACCAATGATGCGACGTGGCCCTTCCGGCCCCAGGACTCTTGCAATGCC
TGTGGACTTTTTTGGGCTAACAGGGGTACATTGAGGGATCTCTCAAAGAAAACAGAAGAG
AATCAGTTGGCTTTAATGAAACCGGATGATGGTGGGAGTGTGCTGATGCTGCTAACAAC
TTAAACACTGAAGCTGCAAGTGTGAAGAACACACTTCCATGGTTTCTCTTGCCAATGGG
GATAATTCTAATCTGTTAGGTGATCACTAA

>G2085 Amino Acid Sequence (domain in AA coordinates: TBD)

MFGRHSIIPNNQIGTASASAGEDHVSASATSGHIPYDDMBEIPHPDSIYGAASDLIPDGS
QLVAHRSDGSELLVSRPPEGANQLTISFRGQVYVFDVAGADKVDVLSLLGGSTELAPGP
QVMELAQQQNHPVVEYQSRCSLPQRAQSLDRFRKKRNARCFEKKVRYGVRQEVRLMAR
NKGQFTSSKMTDGAYNSTGTDQDSAQDDAHPEISCTHCGISSKCTPMMRRGPGSPRTLCA
CGLFWANRGTLRLDLSKKTEENQLALMKPDDGGSVADAANNLNTEAASVEHTSMVSLANG
DNSNLLGDH*

>G1783 (1..603)

ATGGCCGCGTTTCCGCGAGTGGACAAGGGTCGATGACAAACGTTTTGAGTTAGCTCTGCTT
CAAATCCCGGAGGGTTCGCCGAATTTATAGAGAAATATCGCCTATTATCTCCAGAAACCG
GTGAAGGAGGTGGAGTACTACTGCGCGTTGGTCCATGATATTGAGCGGATCGAATCG
GGTAAGTATGTTTTGCCCAAATACCCGGAAGACGATTACGTGAACTGACGGAAGCAGGT
GAGTCTAAGGGCAATGGGAAAAAGACGGGAATTCCTTGGTCAGAAGAGGAACAGAGGTTG
TTTCTGGAAGGACTAAATAAGTTTGGGAAAGGAGACTGGAAGAACATATCGAGGTATTGT
GTGAAGTCAAGGACCTCGACGCAAGTGGCAAGCCATGCTCAGAAGTATTTTGCAAGGCAA
AAGCAGGAGAGTACGAATACTAAACGCCCGAGTATTCATGACATGACTCTGGGAGTTGCG
GTCAATGTCCCTGGATCCAATTTGGAGTCTACTGGCCAGCAACCACATTTTGGTGATCAA
ATTCCTTCGAATCAATATTATCCCTCCAGGAAACTTTTCGGGGTTTTGATCAGCGATGG
TGA

>G1783 Amino Acid Sequence (domain in AA coordinates: 81..129)

MAAFPQWTRVDDKRFELALLQIPEGSPNFIENIAYYLQKPVKEVEYYYCALVHDIERIES
GKYVLPKYPEDDYVKLTAGESKNGKKTGIPWSEEBEQLFLEGLNKFKGDKWKNISRYC
VKSRTSTQVASHAQKYFARQKQESTNTRPSIHDMTLGAVNVPGSNLESTGQQPHFGDQ
IPSNQYYPSQENFRGFDQRW*

>G2072 (155..793)

TCGACCCACGCGTCCGCCACGCGTCCGGATCTTTTCACAGAAGACCAACCAGCTTGGCT
CGATGAGCTCCTAAGTGAGCCAGCATCACCTAAGATTAACAAAGGTCATAGACGTTTCAGC
TAGTGACACAGCTGCTTACTTGAACCTCAGCTTTAATGCCTTCGAAGGAAAATCATGTTGC
TGGTTTCGCTTGGCAGTTCAGAACTATGATTTGTGGCAGTCCAACCTTATGAACAACA
CAATAAATTAGGATGGGATTTCTCTACAGCAAATGGAACATAATATCCAAAGAAATATGTC

ATGCGGAGCTTTAAATATGTCGTCGAAACCCATTGAGAAACATGTAAGCAAAATGAAAGA
 AGGAACCTTACAAAACCAGATGGTCTAGATCAAAGACTGACTCAAAACGTATCAAACA
 TCAAAATGCTCATCGAGCGCGTTTGAGAAGGCTTGAGTACATATCAGACCTTGAAAGGAC
 CATCCAAGTGCTACAAGTTGAAGGATGTGAAATGTCATCTGCCATTCACTACTTGGATCA
 GCAGTTACTCATGCTTAGCATGGAAAAATAGAGCTTTAAACAACGTATGGATAGTTTAGC
 AGAAATCCAAAAGCTTAAACATGTGGAGCAGCAATTGCTTGAGAGAGAGATAGGAAACCT
 ACAGTTTCGACGACACCAACAACAACCACAGCAAAACCAAAAACAAGTCCAAGCAATACA
 AAATCGATACACCAATATCAACCACCTGTTACACAAGAACCCGATGCCCAATTTGCAGC
 CTTGGCAATATGATTTAGGAAATATGGATACATTGTTTCAGATTAAGCTGAGCTCCTCTTG
 CTCTACCTTAATGTCCATACAACATAGGTGAACCTTGATGTTTGTAGCCTTGAATGAAAC
 CTA AAAAAGCATCGTTATGTAAATCAAATGTGGTTGCCCATATCCTCCTCTATTGCATT
 TCTCTCTATTATGGCATGGTAGAGAACTCTTGTCAAGAACTTCATGTTATGTAATAA
 CTTGTAATCCTTCTTATTTTATCTATTATATATGAATAAGTAATTTTTTTGCCAAAAA
 AAAAAAAAAAAAAAAAAA

>G2072 Amino Acid Sequence (conserved domain in AA coordinates: 90-149)

MPSKENHVAGSSWQFQNYDLWQNSYEQHNKLGWDFSTANGTNIQRNMSCGALNMSSKPI
 EKHVSKMKEGSTKPDGPRSKTDSKRIKHQNAHRARLRREYISDLERTIQVLQVEGCEM
 SSAIHYLDQQLMLSMENRALKQRMDSLAETQKLKHVEQQLLEIREIGNLQFRRHQQQPQQ
 NQKQVQAIQNRYTKYQPPVTQEPDAQFAALAI*

>G931 (85..1071)

GGAGGTTCTTTGACAGACACATGTATCATCAATCTTCTCTGTTGAAGCAGAGAGAGAGAG
 AGCTAATTGTTGCCTCTGAGTCACATGGATAAGAAAGTTTCATTTACTAGCTCTGTGGCA
 CATTCAACTCCACCATACTTTAGTACTTCCATCTCATGGGACTTCCAACCAATCCAAT
 GGTGTGACTGAATCACTGAGTTTGAAGGTGGTAGATGCAAGACCAGAACGTCTTATAAAC
 ACAAGAATATCAGTTTCCAGGACCAGGATTCATCTTCAACTCTGTCTCTGTCTCAATCT
 TCTAACGATGTTACAAGTAGTGGAGATGATAACCCCTCAAGACAAATCTCATTTTATAGCA
 CATTGAGATGTTGTAAAGGATTTGAAGAACTCAAAGGAAGCGATTTGCAATTAAATCA
 GGCTCCTCCACGGCAGGAATCGCTGATATTCATCTTCTCCTTCCAAGGCTAACTTCTCA
 TTTCACTATGCCGATCCACATTTTGGTGGTTTAAATGCCCTGCGGCTTACCTACCACAGGCA
 ACAATATGGAATCCCCAAATGACTCGAGTTCCGCTACCATTTCGATCTCATAGAGAATGAG
 CCTGTCTTTGTCAATGCAAAGCAATTCCATGCAATTATGAGGAGGAGGCAACAGCGTGCT
 AAGCTAGAGGCGCAAAACAACATAATCAAAGCCCGTAAGCCGATCTTCATGAATCTCGA
 CATGTTACAGCTCTTAAACGACCTAGAGGATCTGGTGGAGATTCTTAACACCAAAAAAG
 CTTCAAGAATCTACAGATCCAAAACAAGACATGCCAATCCAACAGCAACACGCAACGGGA
 AACATGTCAAGATTTGTGCTTTATCAGTTGCAGAACAGCAATGACTGTGATTGTTCAACC
 ACTTCTCGCTCTGACATCACATCTGCTTCTGACAGCGTTAATCTCTTTGGACACTCTGAA
 TTTCTGATATCAGATTGCCCATCTCAGACAAACCAACATGTATGTTTCATGGTCAATCA
 AATGACATGCATGGAGGTAGGAACACACACCATTTCTCTGTCCATATCTGAGCCGGTGGA
 ATCTGGTAATGTGTACGTTTCTACAAAAAAGGGAAGTCATCCTTGGCTGCTACTTTCGCT
 TATTAGCTAGTTCTTATTTTACACGCTTTGTCCAGATATC

>G931 Amino Acid Sequence (domain in AA coordinates: TBD)

MDKKVSFTSSVAHSTPPYLSISWGLPTKSNVTESSLKVVDDARPERLINTKNISFQD
 QDSSSTLSSAQSSNDVTSSGDDNPSRQISFLAHSVDCKGFEETQQRKFAIKSGSSTAGIA
 DIHSSPSKANFSFHYADPHFGGLMPAAYLPQATIWNPMQTRVPLPFDLIENEPVFNKQ
 FHAIMRRRQRAKLEAQNKLKARKPYLHESRHVHALKPRGSGGRFLNKKLQESTDPK
 QDMPQQQHAATGNMSRFVLYQLQNSNDCDCSTTSRSDITSASDSVNLFGHSEFLISDCPS
 QTNPTMYVHGQSNMDMHGGRNTHHFSVHI*

>G278 (93..1874)

TCGATCTTTAACCAATCCAGTTGATAAGGTCTCTTCGTTGATTAGCAGAGATCTCTTTA
 ATTTGTGAATTTCAATTCATCGGAACCTGTTGATGGACACCATTGATGGATTCGCCC
 ATTCTTATGAAATCAGCAGCACTAGTTTCGTCGCTACCGATAACACCGACTCCTCTATTG
 TTTATCTGGCCGCCGAACAAGTACTCACCGGACCTGATGTATCTGCTCTGCAATTGCTCT
 CCAACAGCTTCGAATCCGTCTTTGACTCGCCGGATGATTTCTACAGCGACGCTAAGCTTG
 TTCTCTCCGACGGCCGGGAAGTTTCTTCCACCGGTGCGTTTGTGTCAGCGAGAAGCTCTT
 TCTTCAAGAGCGCTTTAGCCGCCGCTAAGAAGGAGAAAGACTCCAACAACACCGCCGCCG
 TGAAGCTCGAGCTTAAGGAGATTGCCAAGGATTACGAAGTCGGTTTCGATTTCGGTTGTGA
 CTGTTTTGGCTTATGTTTACAGCAGCAGAGTGAGACCGCCGCTAAAGGAGTTTCTGAAT

GCGCAGACGAGAATTGCTGCCACGTGGCTTGCCGCCGGCGGTGGATTTCATGTTGGAGG
 TTCTCTATTTGGCTTTTCATCTTCAAGATCCCTGAATTAATTACTCTCTATCAGAGGCACT
 TATTGGACGTTGTAGACAAAGTTGTTATAGAGGACACATTGGTTATACTCAAGCTTGCTA
 ATATATGTGGTAAAGCTTGTATGAAGCTATTGGATAGATGTAAAGAGATTATTGTCAAGT
 CTAATGTAGATATGGTTAGTCTTGAAGAGTCATTGCCGGAAGAGCTTGTAAAGAGATAA
 TTGATAGACGTAAAGAGCTTGGTTTGGAGGTACCTAAAGTAAAGAAACATGTCTCGAATG
 TACATAAGGCACTTGACTCGGATGATATTGAGTTAGTCAAGTTGCTTTTGAAGAGGATC
 ACACCAATCTAGATGATGCGTGTCTCTTCATTTGCTGTTGCATATTGCAATGTGAAGA
 CCGCAACAGATCTTTTAAACCTTGATCTTGCCGATGTCAACCATAGGAATCCGAGGGGAT
 ATACGGTGCTTCATGTTGCTGCGATGCGGAAGGAGCCACAATTGATACTATCTCTATTGG
 AAAAGGTGCAAGTGCATCAGAAAGCACTTTGGAAGGTAGAACCGCACTCATGATCGCAA
 AACAAGCCACTATGGCGGTGAATGTAATAATATCCCGGAGCAATGCAAGCATTTCTCTCA
 AAGGCCGACTATGTGTAGAAATACTAGAGCAAGAAGACAAACGAGAACAATTCCTAGAG
 ATGTTCTCCCTCTTTTGCAGTGGCGGCCGATGAATTGAAGATGACGCTGCTCGATCTTG
 AAAATAGAGTTGCACTTGCTCAACGTCTTTTCCACGGAAGCACAAGCTGCAATGGAGA
 TCGCCGAAATGAAGGGAACATGTGAGTTCATAGTGAAGTACTAGCCTCGAGCCTGACCGTCTCA
 CTGGTACGAAGAGAACATCACCGGGTGTAAAGATAGCACCTTTCAGAATCCTAGAAGAGC
 ATCAAAGTAGACTAAAAGCGCTTTCTAAAACCGTGGAACCTCGGGAACGATTCTTCCCGC
 GCTGTTCCGGCAGTGCTCGACCAAGATTATGAAGTGTGAGGACTTGACTCAACTGGCTTGCG
 GAGAAGACGACACTGCTGAGAAACGACTACAAAAGAAGCAAAGGTACATGGAATACAAG
 AGACACTAAAGAAGGCCCTTAGTGAGGACAATTTGGAATTAGGAAATTCGTCCTTGACAG
 ATTCGACTTCTTCCACATCGAAATCAACCGGTGGAAGAGGTCTAACCGTAAACTCTCTC
 ATCGTCGTCGTTGAGACTTTCCTCTTAGTGTAATTTTGTCTGTACCATATAATTCTGT
 TTTTCATGATGACTGTAAGTCTTATGTCTATCGTTGGCGTCATATAGTTTCGCTCTTCGT
 TTTGCATCCTGTGTATTATTGCTGCAGGTGTGCTTCAAACAAATGTTGTAAACAATTTGAA
 CCAATGGTATACAGATTTGTAATATATATTTATGTACATCAACAATAAAAAAAAAAAAAA
 AAAA

>G278 Amino Acid Sequence (domain in AA coordinates: 2-593)
 MDTTIDGFADSYEISSTSFVATDNTDSSIVYLAEEQVLTGPDVLSALQLLSNSFBSVFDSP
 DDFYSDAKLVLSDGREVSFHRCVLSARSSFFKSALAAAKKEKDSNNTAAVKLELKEIAKD
 YEVGFDSVVTVLAYVYSSVRPPKGVSECADENCCHVACRPVDFMLEVLYLAFIFKIP
 ELITLYQRHLLDVVDKVVIEDTLVILKLANICGKACMKLLDRCKEIVKSNVDMVLSLEKS
 LPEELVKEIIDRRKELGLEVPKVKHVS NVHKALDSDDIELVKLLKEDHTNLDACALH
 FAVAYCNVKTATDLLKLDLADVNRNPRGYTVLHVAAAMRKEPQLILSLLEKGASASEATL
 EGR TALMIAKQATMAVECNNIPEQCKHSLKGRLCVEILEQEDKREQIPRDVPPSFAVAAD
 ELKMTLLDLENRVALAQR LFPTEAQAAMEIAEMKGTCEFIVTSLEPDRLTGKRTSPGVK
 IAPFRILEEHQSRLKALSKTVELGKRFFPRCSAVLDQIMNCEDLTQLACGEDDTAEKRLQ
 KKQRYMEIQETLKKAFSEDNLELGNSSLTDSTSSSTSKSTGGKRSNRKLSHRRR*

>G2421 (1..630)
 ATGGAGGGTTTCGTCCAAAGGGTTGAGGAAAGGTGCATGGACTGCTGAAGAAGATAGTCTC
 TTGAGGCACTGTATTGGTAAGTATGGAGAAGGCAAATGGCATCAAGTTCTTTAAGAGCT
 GGGCTAAATCGGTGCAGGAAAAGTTGTAGACTAAGATGGTTAAACTATTTGAAGCCAAGT
 ATCAAGAGAGGAAAATTTAGTTCTGATGAAGTTGATCTTCTTCTTCGTCTTCATAAGCTT
 CTAGGAAATAGGTGGTCCCTTGATTGCTGGTTCGATTACCTGGTCCGACCGCTAATGATGTC
 AAGAACTACTGGAACACCCATCTGAGTAAGAAGCATGAACCGTGTGTAAGTAAGATA
 AAAAGGATAAATATTATAACCCCTCCTAATACACCGGCCCAAAAGTTTGTGAAAATAGT
 ATCAGATGTAACAAAGATGATGAGAAAGATGATTTTGTGGATAATTTTATGGTTGGAGAT
 AATATATGTTGGTGGAGCTTTGTCTAGACGAGGGCCAAGAGGTAGATGTGCTGGTTACAGAA
 GCGGCGGCAACAGAAAAGGAGGGCACTTTGGCGTTTGACGTTGAGCAACTTTGGAATTTG
 TTCGATGGAGAGACTGTGATCTTTGATTAGTGTTTATAAACGTTTGTGTTCTCTTGTGTTG
 TGAGGTTTCTCTATTTAATTTAGTATCTATTTTCTAAATTAATAATATCTTATAGTATT
 TTAGGCAAACCTTATGTTTCCGTTTCTGTGCGGCCGCTCTAG

>G2421 Amino Acid Sequence (domain in AA coordinates: 9-110)
 MEGSSKGLRKGAWTAEDSLLRQCIGKYEGEKWHQVPLRAGLNRCRKSRLRLWNLNLYLKPS
 IKRKGFSSDEVLLLLRLHLKLLGNRWSLIAGRLPGRTANDVKNYWNTHLSKKHEPCCKTKI
 KRINIIPTPNTPAQKYCENSITCNKDDEKDDFVDNFMVGDNIWLERLLDEGQEVLDVLVTE
 AAATEKEGTALAFDVEQLWNLFDEGTIVFD*

>G2032 (53..1789)

TCCCTCCCAGAGTAAGAACTTCCATACTTTGCTCTAGATTTCTTGAGAAAAGATGCAGCC
 GATCTTCCATGCGATCCTTAAAAATGACCTTCCAGCTTTTCTAGAGTTGGTAGAAGATAG
 TGAATCGTCTCTGGAGGAGAGAAAACGAGGAAGAACACTTGAACAACACGGTTTTGCACAT
 GGCTGCAAAGTTTGGTCACCGAGAACTCGTCTCCAAGATTATTGAGCTCCGACCTTCCCT
 CGTGTCTTCCCGCAACGCATACAGAAACACACCTTTGCATCTTGCTGCTATCCTTGGAGA
 TGTAACATAGTTATGCAGATGTTAGAGACTGGATTGGAAGTGTGTTCTGCACGCAATAT
 CAACAACCACACACCACTCCACTTGGCTTGCCGTAGCAATTCCATAGAGGCTGCCAGACT
 CATCGCGGAAAAGACACAATCAATTGSCCTCGGTGAACTCATTCTCGCCATATCAAGTGG
 ATCCACTAGTATCGTAGGGACTATACTGGAGAGATTCCCAGACCTAGCTAGGGAAGAAGC
 TTGGGTGGTTGAAGACGGCTCACAATCAACGCTACTGCATCATGCGTGTGATAAGGGAGA
 CTTTGAAGTGAACATATATTGTTAGGGCTCGATCAAGGATTAGAAGAAGCACTTAACCC
 CAATGGTTTATCACCTCTGCATCTTGCGGTCTCAGAGGCTCGGTTGTGATCCTGGAGGA
 GTTCTTGGACAAGGTTCCATTGTCTTTCAGCTCAATCAGCCGCTCGAAAGAGACAGTCTT
 TCATCTCGTGTCTCGAAACAAAATATGGATGCCTTTGTTTTTATGGCAGAGAGTTGGG
 AATTAACAGCCAAATCTTCTACAGCAACCGATGAAAGTGGCAACACTGTCTTACATAT
 TGCTGCATCCGTCTCTTTTGATGCTCCTCTTATACGTTACATTGTTGGTAAGAATATAGT
 AGATATCACGTCCAAGAACAAGATGGGTTTGAAGCTTTTCAACTTCTCCCTCGAGAAGC
 CCAAGACTTTGAGTTGTATCAAGGTGGCTGAGATTGGTACCGAGACTTCACAAGAGCT
 GGATTCTGAGAACAATGTAGAACAACACGAAGGCTCTCAAGAGGTCGAGGTAATACGGTT
 GCTAAGGATTATAGGAATAAACACATCAGAGATAGCAGAGAGAAAGAGAAGCAAGGAACA
 GGAAGTGGAAAGAGGTCGTGAGAACTTGAATATCAGATGCATATAGAAGCATTACAGAA
 TGCAAGAAATACGATTGCTATAGTGGCAGTCTTGATTGCTTCAGTTGCTTATGCCGGTGG
 GATAAACCCCTCCGGGGGGCGTCTACCAAGACGGGCCATGGAGAGGGAAATCCTTAGTGGG
 GAAAACAACGGCGTTTAAAGTCTTTGCGATATGCAACAACATCGCACTGTTACGTCCTT
 GGGCATCGTTATTCTTCTTGTAGCATCATACCTTACAAGAGGAAACCCTTAAAGAGATT
 ATTGGTGGCCACGCATAGGATGATGTGGGTTTCTGTAGGTTTCATGGCGACGGCTTATAT
 AGCGGCGTCTTGGGTGACCATAACCGATTATCATGGAACACAATGGTTATTTCCAGCAAT
 TGTAGCCGTTGCTGGTGGAGCGTTGACCGTACTCTTTTCTATCTCGGAGTTGAGACCAT
 CGGTCAATGGTTTAAAGAAGATGAATCGTGTAGGGGATAATATACCTTCCTTTGCAAGAAC
 CAGTTCAAGATTTAGCCGCTCCCGGAAAATCAGGCTATTTACCTATTAAGAAAAACTGGT
 TTTCTAATTTCCCTGTAACCTGTGTAATTGTGTATGTG

>G2032 Amino Acid Sequence (domain in AA coordinates: entire protein)

MQPIFHAILKNDLPFALELVEDSESSLEERNEEEHLNNTVLHMAAKFGHRELVSIIELR
 PSLVSRNAYRNTPLHLAAILGDVNIVMQMLETGLEVCARNINNHPLHLACRSNSIEA
 ARLIAEKTQSIGLGELILAISSGSTSIVGTILERFPDLAREEAWVEDGSQSTLLHHACD
 KGDFELTTILLGLDQGLEALNPNGLSPLHLAVLRGSVVILEEFLDKVPLSFSSITPSKE
 TVFHAAARNKNMDAFVFMESLGINSQLLQQTDESGNTVLHIAASVSFDAPLIRYIVGK
 NIVDITSKNKMGFEAFQLLPREAQDFELLSRWLRFGTETSQELDSENNVQHEGSQEVFV
 IRLRLRIIGINTSEIAERKRSKEQEVERGRONLEYQMHEALQNARNTIAIVAVLIASVAY
 AGGINPPGGVYQDGPWRGKSLVGKTTAFKVFAICNNIALFTSLGIVILLVSIIPYKRKPL
 KRLLVATHRMWVSVGFMATAYIAASWVTIPIHYGTQWLFPPIAVAVAGGALTVLFFYLGV
 ETIGHWFKMNRVGDNIPIFARTSSDLAVSGKSGYFTY*

>G1396 (83..313)

TCGACCTCGTTTCCTTTCCTCTCTTCTTACCATTAGTACGTTACTGGAGCTGATCTC
 ACGTATATTTGGATCGTAATCATGGACGGCGAAGATTTGCCGGAAGGCGGCTGCTGA
 AGCCAAGGGATTGAACCCGGGATTAATCGTGTCTGCTTGTGTTGGAGGTCGCTTCTTGT
 GTTCTTAATCGCAACTACGTGCTTTACGTTTATGCTCAGAAGAACCCTACCTCCAAGGAA
 GAAGAAGCCCGTTTCCAAAAGAAGCTCAAGCGGGAGAAGCTAAAGCAAGGAGTCCCTGT
 CCCTGGAGAATAAAAGCCAGCTTAAGCTTCTTCACTTGTGCCTCCTTCAAAGCGGTTTT
 TGTTGGTTACCAAAATTTACCTTGCAGGTTTTTTTTCTTCTTTACTTCTGTCATGAGG
 ATTATCTTTGAGGCCT

>G1396 Amino Acid Sequence (domain in AA coordinates: TBD)

MDGEDFAGKAAAEAKGLNPGLIVLLVVGGLLVFLIANVLYVYAQKNLPPRKKKPVSKK
 KLKREKLKQGVVPVPE*

>G619 (382..2748)

ATTTTTTCCAACTGCAAATTTTAGTCTATGTCTGTTCCCTTGTGCTCCCTCTTCTCAGT

ACCTGCAAATGGAGGAAGAAGAAATCCTTCTCTGAAACCCCTTGTTCTCATTGATTCTCTCC
TTCTCTCTCTTCTTCTCTCTCTGTCTCTGATTTCGTTATTCACACTTATGACTCATCTT
TCCCGTCAATAGCTAAGTTTGCCCTCTTCTTTGTGAAATTTAGCTGAAAAAGGAGAGGAAT
TCCGAATTCGTCTCACTTCAAAGCTCGAATTTTGCAACTTTCCTTTGATGGGTTTTACTT
GTTTTGTGTGTAATCTGATTAAAAATAGAAACTTTTTGTTTTCTTCTTGTCTCCTTTTGCT
CTTAAAGAGAAGCTTTTTCAATGGAATTTGACTTGAATACTGAGATTGCGGAGGTGGAA
GAGGAGGAGAATGATGATGTAGGAGTAGGAGTAGGAGGAGGAACAAGAATTGACAAGGGT
AGGCTTGAATTTACCATCTTCTTCTTCTTCTCATGCTCTTCCGGATCATCATCGTCATCA
TCTTCTACAGGCTCTGCATCTTCCATTTACTCTGAGCTTTGGCATGCTTGTGCTGGTCCT
CTCACTTGTCTTCCCAAGAAAGGCAATGTAGTTGTCTATTTCCCTCAAGGTCATTTGGAG
CAAGATGCTATGGTTTCATATTCGTCTCTCTTGAAATCCCCAAATTTGACCTTAATCCC
CAAATCGTCTGCAGGGTGGTTAATGTCCAGTTGCTTGTCTAATAAGGACACCGATGAGGTC
TACACTCAAGTCACTCTGCTTCCACTTCAAGAGTTTTCGATGCTAATAAGGGAGGGGAAA
GAGGTCAAGGAGTTAGGAGGGGAGGAAGAGAGGAACGGAAGCTCATCCGTCAAGCGGACA
CCTCATATGTTCTGTAAAACCTTAACAGCGTCTGACACAAGCACACATGGAGGCTTCTCT
GTACCTAGAAGAGCCGCTGAAGATTGTTTTGCTCCTCTTGACTACAAACAACAGAGGCCA
TCTCAAGAGCTCATTGCAAAGGACCTCCATGGAGTAGAGTGGAAGTTTCGCCATATCTAT
AGAGGTCAACCAAGGAGGCATCTACTCACCCTGCTTGGAGTATCTTTGTCAGTCAAAAG
AATCTCGTCTCTGGTGATGCGGTTCTCTTTCTGAGAGACGAAGGAGGAGAGCTGAGATTA
GGAATCAGAAGAGCAGCACGGCCAGAAATGGACTTCTGACTCAATCATTGAGAAGAAT
TCATGTTCAAACATCTCTGCTCTTGTGGCTAATGCTGTATCTACAAAAGCATGTTTCAT
GTGTTCTACAGTCCACGACGACGCATGCAGAGTTTGTGATTCTTATGAGAAGTATATC
ACAAGCATCAGGAGTCTGTTTGCATAGGCACAAGATTAGAATGCGATTGAAATGGAC
GATTCTCCTGAGAGAAGATGCGCTGGTGAGTACTGGAGTCTGTGACTTGGACCCGTAT
AGGTGGCCAACTCTAAATGGAGGTGCTTGTGGTGCGATGGGATGAGTCTTTGTGAGT
GATCATCAAGAAAGAGTTTACCTTGGGAGATTGATCCCTCGGTTTCTCTCCCACTTG
AGCATTCACTCATCTCCAAGGCCTAAAAGGCCATGGGCAGGTTTACTGGATACTACCCCA
CCCGAAACCCCATAAACAAAAGGGGTGGTTTTTTGGACTTTGAGGAGTCCGTTAGACCC
TCTAAGGTCTTGCAAGGTCAAGAAAATATAGGTTCTGCATCACCTCACAGGGGTTTGAT
GTTATGAACCGCCGATACTGGATTGTGCGATGCAGTCTCATGCAAAATCCAGTCTTTGTG
TCGAGTAGAGTCAAGGATCGATTGTGTGAGTTTGTAGATGCTACTGGCGTGAACCCAGCT
TGTTCAAGGTGTTATGGACCTGGATAGGTTTCCAAGGGTCTTGCAAGGTCAAGAAATTTGC
TCGCTTAAATCATTCCCGCAATTTGCTGGTTTCACTCCAGCTGCTGCTCCTAATCCCTTT
GCTTACCAAGCCAACAAGTCAAGTTACTATCCGCTAGCTTTGCATGGGATTAGGAGCACT
CATGTTCCGTATCAGAATCCATACATGCGGGAAACCAATCCTCGGGTCCCCCTTCACGT
GCAATAAACTTTGGTGAAGAGACTAGAAAAGTTTGATGCACAAAATGAAGGTGGCCTACCA
AATAATGTTACAGCTGATTGTCATTCAAGATTGATATGATGGGAAAACAGAAAGGCAGT
GAGTTGAATATGAATGCTTCAAGGATGTAACTTTTCGGATTCTCCTTACCAGTGGAG
ACACCTGCATCTAAGCCGCAAGCTCGAGCAAAAGAATCTGTACAAAGGTTCAACAAGCAA
GGAAGCCAAGTGGGGAGAGCTATTGATTGTGTCGCACTTAACGGGTATGATGATCTCCTT
ATGGAGCTTGAACGGCTGTTCAACATGGAAGGGCTTCTCAGGGATCCTGAAAAAGGATGG
AGGATCTTATATACTGATAGTGAGAACGATATGATGGTCTTGGCGATGATCCATGGCAT
GATTTCTGCAATGTGGTGTGGAAGATACACTTATACACGAAAGAGGAAGTGGAGAATGCG
AATGACGATAACAAGAGTTGTTTAGAGCAAGCTGCTCTCATGATGGAAGCATCAAAGTCA
TCTTCTGTGAGCCAGCCTGATCTTCTCTCTACAATCACTAGGGTTTGATACCCATAAAGA
AGCTTATTTCTATGTTTAAAGTGTGTTTGTCTCACAAAAGAACTTCACTTTATCTTT
GTCTTTGAATCCATTTATGTGTTTGTGTTTCTTCTGCTCCATGGATGCTCATG
TGTACCGTTTACTGAGAGATATGTGAGTTTATGGGATGTGTAAAGCATGCCATTGGAT
TTTAAGGTTTTCAAAATTACAATATATATATTAGTTTTGAAGTTAAAAA

A
>G619 Amino Acid Sequence (domain in AA coordinates: 64-406)
MEFDLNTIEIAVEVEEENDDVGVGVGGGTRIDKGRLGISPPSSSSSSSSSSSSSSSSSSSSSGSAS
SIYSELWHACAGPLTCLPKKGNVVVYFPQGHLEQDAMVSYSSPLEIPKFDLNPQIVCRVV
NVQLLANKDITDEVYTVQVTLPLQEFSLNNGEGKEVKELGGEEERNNGSSSVKRTPHMFCKT
LTASDSTHGGFSVPRRRAEDCFAPLDYKQQRPSQELIAKDLHGVEWKFRHIYRGQPRRH
LLTTGWSIFVSQKNLVSGDAVLFLRDEGGELRLGIRRAARPRNGLPDSIIIEKNSCSNILS
LVANAVSTKSMFHVFSRATHAEFVIPYEKYITSIRSPVCIGTRFRMRFEMDDSPERRC

AGVVTGVCDDLDPYRWPNKWRCLLVWRWDESFVSDHQERVSPWEIDPSVSLPHLSIQSSPR
 PKRPWAGLLDTPPGNPITKRGGFLDFEESVRPSKVLQGOENIGSASPSQGFDMNRRIL
 DFAMQSHANPVLVSRVKDRFGEFVDATGVNPACSGVMDLDRFPRVLQGOEICSLKSPQ
 FAGFSPAAPNPFAYQANKSSYYPLALHGIRSTHVPYQNPYNAGNQSSGPPSRAINFGEE
 TRKFDAQNEGGLPNNVTADLPFKIDMMGKQKGSELNMNASSGCKLFGFSLPVETPASKPQ
 SSSKRICKVKHKQGSQVGRAIDLRLNGYDDLMELELRFNMEGLLRDPEKGWRILYTD
 ENDMMVVGDDPWHDPCNVVWKIHLTYKKEVENANDDNKSCLEQAALMMEASKSSSVSQPD
 SSTITRV*

>G2295 (33..917)

GTAATATATAACAATAACTCAGGTTACAAAGGATGGTTCCGAAAGTGGTCGACCTACAAA
 GGATAGCGAACGATAAGACAAGGATAACAACCTACAAGAAGAGGAAAGCTAGTCTTTACA
 AGAAGGCACAAGAGTTCTCAACTCTCTGCGGCGTCGAGACATGTCTCATCGTCTACGGTC
 CCACGAAGGCTACCGATGTGGTGATTTCCGAGCCAGAGATATGGCCGAAGGACGAGACCA
 AAGTCAGGGCCATCATACGCAAGTACAAAGACACAGTGTGACCCAGCTGCAGGAAAGAAA
 CCAACGTGGAGACTTTTCGTCAACGATGTAGGGAAAGGAAACGAGGTGGTGACTAAAAAGA
 GAGTGAAGCGTGAGAATAAGTATTCTAGTTGGGAGGAGAAGCTAGACAAGTGTTACGAG
 AGCAACTACATGGGATTTCTGTGCGGTGGATAGCAAGTTAAATGAAGCTGTAACGAGAC
 AGGAGCGTAGTATGTTTAGGGTTAATCATCAAGCCATGGACACACCATTTCCCGCAGAATT
 TAATGGACCAACAATTCTATGCCACAGTATTTTCATGAGCAGCCACAGTTTCAAGGCTTCC
 CTAATAATTTCAATAATATGGGTTTCTCGTTGATTTACCTCATGATGGTCAGATTCAAA
 TGGACCCAAATCTCATGGAGAAGTGGACCGACTTGGCTTTGACTCAAAGCTTGATGATGT
 CAAAGGGAAACGATGGTACTCAATTCATGCAGAGGCAAGAACAACCATACTATAATCGTG
 AACAGGTGTATCGAGGTCTGCAGGTTTCAATGTTAACCCTTTATGGGATATCAAGTCC
 CGTTTAAATATTCCTAATTGGAGATTATCGGGAAATCAAGTTGAAAATTGGGAGCTTTTCAG
 GGAAGAAAACGATATGATTGAATTACGGAGCTTTATTAGTTTTTAGGGTTTTATAGTTT
 TG

>G2295 Amino Acid Sequence (domain in AA coordinates: TBD)

MVPKVVDLQRIANDKTRITTYKKRKASLYKKAQEFSTLCGVETCLIVYGPTKATDVVISE
 PEIWPKDETKVRAIRKYKDTVSTSCRKETNVETFVNDVGKGNVVTKKRVKRENYSSW
 EEKLDKCSREQLHGIFCAVDSKLEAVTRQERSMFRVNHQAMDTFPQNLMDQQFMPQYF
 HEQPQFGFPNNFNMMGFSLISPHDGIQMDPNLMEKWTDLALTQSLMMSKGNDDGTQFMQ
 RQEQPYNREQVVSRSAGFNVNPFMGYQVPFNIIPNWRLSGNQVENWELSGKKTII*

>G312 (1..1755)

ATGGCTTACATGTGCACTGATAGTGGCAATCTAATGGCTATTGCTCAACAAGTCATCAAA
 CAGAAGCAGCAACAAGAACAACAACAGCAGCAACATCATCAAGACCATCAGATTTTTGGT
 ATTAATCCTTTGTCTCTTAACCCATGGCCCAATACTTCCCTCGGGTTTGGGCTTTTCAGGT
 TCGGCTTTTCCCGACCGGTTTCAAGTTACCGGCGGCGGAGATTCCAACGATCCTGGCTTT
 CCTTTTCTTAACCTTAGACCACCAACACGCCACAACCACCGGCGGTGGGTTTCAGGTATCT
 GATTTTCGGCGGTGAACCGGCGGCGGCGAGTTTGAGTCCGACGAGTGGATGGAGACTCTT
 ATCAGCGGTGGAGACTCCGTTGCAGACGGTCCCTGATTGTGACACCTGGCATGATAATCCC
 GATTACGTAATCTACGGTCCCTGATCCATTTCGATACTTACCCGAGTCGACTCAGTGTCCAA
 CCGTCAGATCTAAACCGAGTCATTGACACGTCGAGTCCGCTTCTCCGCCGACCTTGTGG
 CCTCCTTCTTCGCCATTATCGATTCTCCTCGCTTACTCATGAGTCACCAACCAAGAAGAT
 CCAGAGACTAACGACTCCGAAGACGATGACTTCGACCTAGAACCACCTCTCCTCAAAGCT
 ATATACGACTGTGCACGGATCTCAGACTCTGACCCTAACGAAGCTTCCAAGACGCTTCTT
 CAGATCCGAGAATCTGTATCGGAGCTAGGTGATCCGACGGAGCGAGTTGCATTTTACTTC
 ACGGAAGCTCTCTCAACAGACTGTCTCCTAATTCGCCGCGGACGTCGTCTTCTTCTTCA
 TCTACGGAGGATTTAATCTTATCTTATAAAACCCTAAACGACGCTTGTCTTACTCCAAA
 TTCGCACATTTGACGGCGAATCAAGCGATTCTAGAAGCGACGGAGAAGTCGAACAAGATT
 CACATCGTCGATTTTGAATCGTTCAAGGTATACAATGGCCTGCTCTTCTTCAAGCTCTA
 GCTACTCGTACTTCTGGTAAACCCACTCAAATCCGGGTCTCGGGTATACCCGCTCCATCT
 CTCGGTGAATCTCCGGAACCGTCTGTTAATCGCCACCGGAAACCGCTCCGTGATTTCCGC
 AAGGTTCTGGATCTGAATTTGATTTTCATCCCAATTCCTACTCCCATACATTTACTTAAC
 GGGTCAAGTTTCCGGGTGACCCGGATGAAGTACTGGCCGTGAATTTTCATGCTCCAGCTC
 TACAAATTACTCGACGAGACGCCGACGATAGTTGACACCGCACTACGGCTCGCCAAATCG
 TTGAACCCGAGGGTCTGCTACTCTCGGAGAATACGAAGTGAGCTTAAACCGGGTCCGTTTC
 GCTAACCCGGTAAAGAACCGCTTCAATTCTATTCCGCGTTTTTCGAATCCCTTGAACCG

AACTTGGGGCGTGATTTCGGAGGAGAGAGTGAGAGTTGAGCGAGAGTTGTTTCGGCCGGAGA
 ATCTCGGGTTTGATTGGACCGGAGAAAACCGGAATTCATAGAGAAAGAATGGAAGAGAAA
 GAGCAATGGCGGGTATTAATGGAGAATGCCGGTTTTGAATCGGTTAAGCTGAGTAATTAC
 GCAGTGAGCCAAGCAAGATTCTATTGTGGAATTACAATTACAGCAATTTGTATTCAATT
 GTTGAATCTAAGCCTGGCTTCATCTCTTTGGCCTGGAACGATTACCTCTCCTCACTCTT
 TCTTCTGGCGATAA

>G312 Amino Acid Sequence (domain in AA coordinates: 320-336)
 MAYMCTDSGNLMAIAQQVIKQKQQEQQQQHHQDHQIFGINPLSLNPWPNTSLGFGLSG
 SAFPDPFQVTGGGDSNDPGFPFNLDDHHHATTTGGGFRLSDFGGGTGGGEFESDEWMETL
 ISGGDSVADGPDCTDTHDNDPDYVIYGPDPFDTPSRLSVQPSDLNRVIDTSSPLPPPTLW
 PPSSPLSIPPLTHESPTKEDPETNDSDDDDFDLEPPLLKAIYDCARISDSDPNEASKTLL
 QIRESVSELGDPTEVAFYFTEALSNNLSPNSPATSSSSSSSTEDLILSYKTLNDACPYSK
 FAHLTANQAILEATEKSNKIHIVDFGIVQGIQWPALLQALATRTSGKPTQIRVSGIPAPS
 LGESPEPSLIATGNRLRDFAKVLDLNFDFIPILPTIHLNNGSSFRVDPDEVLA VNFMLQL
 YKLLDEPTIVDTALRLAKSLNPRVVTLGEYEVSLNRVGFANRVKNALQFYSAVFESLEP
 NLGRDSEERVVRERELFGRRISGLIGPEKTGIHRERMEEKEQWRVLMENAGFESVKLSNY
 AVSQAKILLWNYNYSNLYSIVESKPGFISLAWNDLPLLLTLSSWR*

>G1444 (192..1001)

AATCCCCTATCCTTCGCAAGACCCTTCCTCTATATAAGGAAGTTCATTTTCTTTT
 GACACGCTGACAAGCTGACTCTAGCATATCTGGCACC GGCGACAGTCTCTTTGGTGC
 AAAGATCCCAAAAATCAAATCGAAAGAGAGAATAAATCAAAAGGAAGAATCTTTATCT
 GCTTTCTCTCGATGAGGATCCGGAACGACAAGTGCCTCTTCCTTTATCGTCTCTATTAC
 CAGTTCTCTATCAGATCTCTACTTTAACCGCTCACCGACGGCCACCGCAGATACTTTC
 GCGGTGGTTATAAAGACGGCGGTGATGATTTTGGTTCTCTTCAGCTTTCGCTTCGCCGC
 CGTCGCAGATTTCTGATCGGCTTATTCAAAGAGATTTGATAAAGAAGAAGGAGGAGTCA
 AGCCTTTGGATGATGATAATGGTGATGTAGACGTCAAGAGTCGTACTGATGCATCGGGCA
 GCAAGAATGTTAATCCCCGAGGAGAATCCGTCTCTTCAATACAAGTTGTGAGAAGAATG
 AAAAGGTTGTGTCTTTGAGGAAGAGAAGAGGCTTTATCAACTTTGAGGATTACGAAGATG
 AGGAAGATGAAGAAGCTAGTGGCGGTGGAGGCCGTATTAATAAAGGGAAAAAGAAAGCGA
 AAAAGAGCGGTGGTGGGTAGAGGAAGGATCACGGTGCAGCCGTGTTAACGGTAGAGGAT
 GGAGATGTTGTGACGAAACGCTTGTGGTTATTCTCTTTGTGAGCATCATCTCGGTAAAG
 GAAGGGTAAGGAGCATGAACAAGAGTGGTGGTGGTGGTGGTGGCGGAGAAAGAGCGGTGG
 TGGTGAAGTGAAGAAGAAGAGAGTAAAGCTTGGCATGGTAAAGGCACGTTCAATAAGTA
 GTTTGCTTGGACAAACCAGCACTAGTGGTGGTACTAGTGGTGATGTTGATCAGGGTGAGA
 TAAGTGCACCTGCTGATCAGTTGCTGATGTGATAAGTAGGTCTGTTGATCAGCATTTG
 CATGTATATGGATATGTGTATGTTTATGTATCATGATGATAATGGGCATAGCGCGCCGCT
 CTAGACAGGCCCTGGAACCGGATCCTCTAGCTAGAGCTTTCTGTTAGTATCATCGGGTTAG
 ACAACGTT

>G1444 Amino Acid Sequence (domain in AA coordinates: 168-193)
 MRIRKRVPLPLSSLLPVPLSDLYFNRSPTATARYFRGGYKDGDDDFGSLQLSLPPPSQI
 SDRLIQRDLIKKKEEVKALDDNDGDVDVKSRTDASGSKNVNPRGESVSSIQVVEKNEKVV
 SLRKRRGFINFEDYEDEEDEEASGGGRINKGKKKAKKSGGGLEEGSRCSRVNRRGWRCC
 QQTLVGYSLCEHHLGKGRVRSNMKSGGGRGGEKKAVVVEVKKRVKLGVMKARSISLLG
 QTSTSGGTSVDQGEISAPADQFAACDK*

>G801 (27..746)

GATAGTGATAACGAAATCCTAATTCATGGCCGACAACGACGGAGCAGTGAGTAACGGCA
 TCATAGTCGAGCAGACGTCAAACAAAGGACCTCTTAACGCCGTTAAGAAACCACCGTCTA
 AAGATCGACACAGCAAAGTTGACGGAAGAGGAAGAAGGATTCGTATGCCAATCATTTGCG
 CAGCTCGAGTTTTTCAATTGACCAGAGAGTTAGGTCAAGTCCGATGGTCAAACCATAG
 AGTGGCTTCTCCGTCAAGCTGAGCCTTCTATCATAGCCGCCACTGGAACCTGGCACTACTC
 CGGCGAGTTTCTCCACTGCTTCTCTCTCCACTTCTTCTCCGTTTACTCTCGGGAAACGTG
 TCGTCAGAGCGGAGGAAGGAGAATCCGGCGGCGGAGGAGGAGGGTTAACAGTGGGAC
 ACACAATGGGGACTTCGTTAATGGGTGGTGGTGGTTCGGTGGGTTTTGGGCTGTTCCGG
 CGAGGCCGATTTCCGACAAGTCTGGAGCTTTGCAACCGGAGCTCCACCGGAAATGGTTT
 TTGCGCAGCAGCAGCAACGCTACACTCTTCGTCGCCACAGCAGCAACAGCAAGCTT
 CCGCCGCCGAGCAGCTGCAATGGGTGAGGCTTCAGCAGCTAGAGTTGGGAATTATCTTC
 CGGGTCATCATCTCAATTTGCTTCTTTGTCTGGTGGAGCTAACGGGTGGGTTCGGA

GGGAAGACGACCACGAACCACGTTGAGAAATGGTATTGTCTTTTGGTAATGTATAGAAA
AATTCCTATGTTTATGTCATCGAAAGTGTTTAGAAAGTACCTCTAATTTGCGGTTTCTT
TTGCTCCTTTTACTTAATTTAAGCTTATTGCTTGTTGATTAGGGTTTTAGGGTTTAA
GAATATTTGGTCTCGTTAATTTGTTTCGGAGAGTGATAGAAAAGAGAGAGATTGATTGA
TTGTTGTACCTAAAACGCTATAAAAGCTCTGTTTTACTAGCGAAAAAA

>G801 Amino Acid Sequence (domain in AA coordinates: 32-93)
MADNDGAVSNGIIVEQTSNKGPLNAVKKPPSKDRHSKVDGRGRRIRMPIICAARVFQLTR
ELGHKSDGQTIEWLLRQAEPSIIAATGTGTPASFSTASLSTSSPFTLGKRVVRAEEGES
GGGGGGGLTVGHMTGSLMGGGGSGGFVAVPARPDFGQVWSFATGAPPEMVFAQQQQPAT
LFVRHQQQQASAAAAAAMGEASAARVGNYPGHHLLLASLSGGANGSGRREDDHEPR*
>G1950 (42..764)

CTGAATTCGAACTTTGGAAGAAGAAGAAGCTTTGATCAATCATGGAAATTGCAACCGATA
CAGCAAAGCAGATGAGAGACGAAGAGTTGTTCAAAGCAGCGGAATGGGGAGATTTCATCGT
TGTTTCATGTCATTATCTGAAGAACAGCTCTCTAAATCTCTCAATTTAGAAAACGAAGATG
GTCGCTCTCTCCTCCATGTGCTGCTTCCTTCGGCCATTCTCAAATAGTGAAGTTGTTAT
CAAGTTTCAGATGAAGCAAAGACTGTAATCAATAGCAAGGATGATGAAGGATGGGCTCCTT
TGCATTCCGCTGCTAGCATCGGTAATGCTGAGCTCGTTGAGGTGCTTTTGACCAGAGGTG
CTGATGTCAATGCCAAAAATAACGGTGGTCCGACTGCTCTTCACTATGCTGCTAGCAAAG
GCCGGTTGGAGATTGCTCAGCTTTTATTAACACACGGTGCAAAGATTAACATCACAGACA
AGGTTGGTTGCACTCCGCTTCACAGGGCAGCAAGCGTGGGAAAGTTAGAAGTTTGTGAAT
TTCTTATTGAAGAAGGAGCAGAGATCGATGCTACGGATAAAATGGGTCAAACCTGCACTCA
TGCATTCACTTATCTGCGATGACAAACAGGTTGCGTTCCCTGCTTATAAGACATGGTGCAG
ATGTGGATGTAGAAGACAAGGAAGGCTACACTGTTCTAGGCCGAGCTACCAATGAATTCC
GACCTGCACCTTATCGATGCTGCTAAGGCCATGCTTGAAGGATAAAATGACTCTGGATTAC
TTTAAACTTACTAATCTGAGAGTTGTTTAGTTACTTAAAGGATTTTCTTTACTGTA
TCATGTTTGCAAATGTTTCTGCCTTATCAATTCATGTTCTGT

>G1950 Amino Acid Sequence (domain in AA coordinates: 65-228)
MEIATDTAKQMRDEELFKAAEWGDSSLFMSLSEQLSKSLNFRNEDGRSLLHVAASFGHS
QIVKLLSSSDEAKTVINSKDDEGWAPLHSAASIGNAELVEVLLTRGADVNAKNNGGRTAL
HYAASKGRLEIAQLLLTHGAKINITDKVGCTPLHRAASVGKLEVCEFLIEEGAEIDATDK
MGQTALMHSVICDDKQVAFLLIRHGADVDEDKEGYTVLGRATNEFRPALIDAAKAMLEG
*

>G958 (55..1950)
CGTCGACATGTTTCATTTTGTCTTAGCTAAGAAGTTTGTATAAGGCAGTGGACATGGCT
CCTGTTTCAATGCCTCCAGGTTTCCGGTTTCATCCAACAGACGAAGAGCTTGTATATAC
TACCTCAAGCGAAAGATTAATGGTCCGACTATTGAGTTAGAGATAATACCCGAGATTGAT
CTTTACAAATGCGAACCTTGGGATTTACCTGGGAAGTCCTTGCTGCCAAGTAAAGACCTA
GAATGTTCTTTTTTTCAGTCTCGAGACCGGAAATATCCAAACGGATCAAGAACAAACCGG
GCGACCAAAGCAGGTTACTGGAAAGCCACCGGGAAAGATCGTAAAGTGACTTCACATTCA
CGGATGGTTGGAACAAAGAAAACATTAGTTTATTACCGAGGAAGAGCGCTCATGGCTCT
CGTACCGATTGGGTTCATGCACGAGTACCGTCTTGAAGAACAAGAATGTGACTCTAAATCC
GGTATACAGGATGCCATGCACTTTGTGCGAGTATTTAAGAAGAGTGCTTTAGCCAACAAA
ATTGAAGAACAACACCATGGTACGAAGAAGAACAAGGAACGACTAATAGTGAACAATCT
ACTTCTAGTACTTGTGTTGATTCTGATGGAATGTATGAAAACCTCGAAAACCTCGGGGTAT
CCAGTCTCACCTGAGACAGGAGGCTTAACCTCAACTCGGTAATAATTTCGTCGTCGGATATG
GAAACGATAGAGAATAAATGGAGTCAGTTTATGTCGCATGACACGTCCTTCAACTTCCCA
CCTCAGTCTCAATATGGAACAATCTCATATCCTCCCTCGAAGGTTGATATAGCGTTAGAG
TGTGCAAGACTACAAATCGTATGTTGCCACCAGTACCACCCTTTACGTAGAAGGTCTC
ACACACAATGAATATTTTGGAAACAATGTAGCTAACGATACAGATGAAATGTTGAGCAAG
ATTATAGCATTGGCTCAAGCCTCACATGAGCCACGAAACAGTCTAGACTCATGGGACGGT
GGTTCTGCTTCCGGGAACCTTCATGGAGACTTTAATAATTCCGGAGAAAAAGTCTCATGC
CTAGAGGCGAACGTGGAGGCTGTAGATATGCAAGAACACCATGTGAATTTTAAAGGAAGAA
AGACTTGTGAAAACCTTGAGATGGGTAGGAGTATCAAGCAAGGAACCTGAAAAGAGCTTC
GTTGAAGAACACTCAACGGTAATTCCTATAGAAGATATTTGGAGATATCATAATGATAAT
CAAGAACAAGAACATCATGATCAAGATGGTATGGACGTTAACAACAACAATGGAGATGTG
GATGATGCTTTTCACTCGAGTTTTCGGAAAACGAACATAACGAGAATCTTTTGGACAAG
AACGATCATGAGACAACGAGTTCCTCATGTTTTGAGGTGGTAAAAAAGTTGAGGTTAGC

CATGGATTGTTTGTACAACTCGTCAGGTAACCAACACATTCTTCCAACAGATAGTACCA
TCGCAAACCGTTATAGTTTATATAAATCCGACGGATGGCAATGAGTGTTCATAGTATG
ACATCAAAAGAGGAGTTTCATGTCCGTAAAAAGATAAATCCGCGAATCAACGGAGTAAGC
TCAACAGTTCTTGGACAATGGAGAAAATTCGCGCATGTTATTGGCTTCATTCCTATGCTT
CTATTGATGCGTTGTGTTTCATCGAGGTAACCTAACAAAAACAGAGGCAGTGAAGGTTAC
TCGAGGCAGCCTACGAGAGGAGATTGTAACAATCGGGGAACAATACTCATGATGGAAAAT
GCTGTCGTGAGAAGAAAAATTTGGAAGAAGAAGAAAGAGAAAAATATGGTTGACGAACAA
GGTTTTCGGTTTCAAGATAGTTTCGTATTGAAGAAGTTGGGGCTTTCTCTTGCTATCATC
TTAGCTGTTTCTACCATAAGTCTTATTTGAATACTGAGGTTCAATATATCATATATGGCT
TTTCACTTTTCTATTGTACTCCCATTTGCCTAGGTCGTATGC

>G958 Amino Acid Sequence (conserved domain in AA coordinates: 7-156)
MAPVSMPPGFRFHPDEELVIYYLKRKINGRTIELEIIPIDLYKCEPWLDPGKSLLP
DLEWFFFSRDRKYPNGSRNTNRATKAGYWKATGKDRKVTSHSRMVGTCKTLVYYRGRAPH
GSRTDWMHEYRLEBQBCDSKSGIQDAYALCRVFKKSALANKIEEQHHGTTKKNKGTNNSE
QSTSSTCLYSDGMYENLENSGYVSPETGGLTQLGNSSSDMETIENKWSQFMSHDTSPN
FPQSQYGTISYPPSKVDIALECARLQNRMLPPVPLYVEGLTHNEYFGNNVANDTDEML
SKIIALAQASHEPRNSLDSDWGGASGNFHGDFNYSGEKVSLEANVEAVDMQEHVNF
EERLVENLRWVGVSSEKELEKSFVEEHSVIPIEDIWRHYHNDNQEQEHDQDGMVNNNG
DVDDAFTLEFSENEHNENLLDKNDEHTSSSCFEVVKKEVSHGLFVTTTQVNTNFFQOI
VPSQTVIVYINPTDGNCCSHMSTKEEVHVRKKINPRINGVSVTLGQWRKFAHVIGFIP
MLLLMRCVHRGNSNKNRGSEGYRQPTRGDCNNRGITILMMENAVVRRKIWKKKKEKNMVD
EQGFRFQDSFVLKLLGLSLAILAVSTISLI*
>G1037 (1..1722)

ATGACTGTTGAACAAAATTTAGAAGCTTTGGATCAGTTTCCTGTAGGAATGAGAGTTCTT
GCTGTTGATGATGACCAAACCTTGTCTCAAAATCCTTGAATCTCTCCTTCGTCACTGCCAA
TACCATGTAAACAACGACGAACCAAGCACAAAAGGCTTTAGAGTTATTGAGAGAGAACAAAG
AACAAGTTTGATCTGGTTATTAGTGATGTTGACATGCCTGACATGGATGGTTTCAAACCT
CTTGAGCTTGTGGTCTTGAAATGGACCTACCTGTCTAATGTTGTCTGCGCATAGTGAT
CCAAAGTATGTGATGAAGGAGTTACTCATGGTGCTTGTGATTATCTACTGAAGCCGGTT
CGTATTGAGGAGTTGAAGAACATATGGCAACATGTCGTGAGAAGTAGATTGATAAGAAC
CGTGGGAGTAATAATAATGGTGATAAGAGAGATGGATCAGGTAATGAAGGTGTTGGGAAT
TCTGATCCGAACAATGGGAAAGGTAATAGAAAACGTAAAGATCAGTATAATGAAGATGAG
GATGAGGATAGAGATGATAATGATGATTCTGTGCTCAAAAGAAGCAACGTGTTGTTTGG
ACTGTTGAGCTGCATAAGAAATTTGTTGCAGCTGTTAACCAATTGGGATATGAGAAGGCT
ATGCCTAAAAGATTTTGGATCTGATGAATGTTGAGAAGCTCACTAGAGAAAATGTGGCC
AGTCATCTTCAGAAATTCGCCTTTACTTGAAGAGGATCAGTGGTGTGGCTAATCAGCAA
GCTATTATGGCAAACTCTGAGTTACATTTTATGCAAATGAATGGACTTGATGGTTTCCAT
CACCGCCCAATCCCTGTTGGATCTGGTCAGTACCATGGTGGGGCTCTGCAATGAGATCT
TTCCCTCAAACGGGATTCTTGGCAGACTCAATAGCTCTTCGGGGATCGGTGTCCGCAGC
CTTTCTTCTCCTCTGAGGAATGTTCTTGCAAAACAGACCGATATCGGAAAGTTTCAC
CATGTCTCATCACTTCTCTTAACCACAGTGATGGAGGAAACATCTTCAAGGGTTGCCA
ATGCCTTTAGAGTTGACACAGCTTCAGACAAACAACAACAAAAGTAGAAACATGAACAGT
AACAAGAGCATTTGCTGGGACCTCCATGGCTTTTCTAGCTTCTCTACGCAACAAAACCTCG
CTCATCAGTGCTCCTAATAACAATGTCGTGGTTCTAGAAGGTCACCCACAAGCAACTCCT
CCAGGCTTCCAGGACACCAGATCAATAACGTTTGGAGCATTTGGTCAAATGCTGTATCC
TCTTCGACTCAACCTCCTCCCCGGCACATAACAGTAATAGTATCAATCATCAGTTTCGAT
GTCTCTCCATTACCGCATTCTTAGACCCGACCCCTTGGAAATGGAACAATGTGTATCAAGC
TACTCTATACCATTCTGTGACTCTGCCAATAACATTGAGTTCTCCAGCCTTGGATACAACA
AATCCCGAGCTTTCTGTAGAAACACGGACTTCGATTCAAACACAAATGTGCAACCTGGA
GTCCTTTATGGTCCATCCACGGATGCTATGGCTCTGTTGAGTAGTAGTAACCCGAAAGAA
GGGTTCCGTAGGCCAACAGAGTTACAGAGTGGTGGATTCTAGGTTGCAGATGCTGGT
TCCTTAGATGATATAGTCAACTCCACGATGAAGCAGGTGTA

>G1037 Amino Acid Sequence (domain in AA coordinates: 11-134, 200-248)
MTVEQNLEALDQFPVGMRLAVDDDDQCLKILESLLRHCQYHVTNNAQKALELLRENK
NKFDLVISDVMDPMDGFKLLELVGLEMDLPVIMLSAHSDPKYVMKGVTHGACDYLLKPV
RIEELKNIWQHVVRSRFDKNRGSNNNGDKRDGSGNEGVSNDPNNKGKRNKRKDQYNEDE
DEDRDDNDDSCAQKKQVRVVTVELHKKFVAAVNQLGYEKAMPKKILDLMNVEKLTRENV

SHLQKFRLLYLRISGVANQQAIMANSELHFMQMNGLDGFHHRPIPVGSGQYHGGAPAMRS
PPNGILGRNLSSSSGIGVRSLSPPAGMFLQNQTDIGKFHHVSSLPLNHSDDGNNILQGLP
MPLEFDQLQTNNNKSRNMNSNKS IAGTSMAPFSFSTQQNSLISAPNNNVVLEGHPOATP
PGFPGHQINKRLEHWSNAVSSSTHPPPAHNSNSINHQFDVSPHPSRPDPLEWNNVSSS
YSIPFCDSSANTLSSPALDTTNPRAFRCNTDFDSNTNVQPGVFYGPSTDAMALLSSSNPKE
GFVVGQQKLQSGGFMVADAGSLDDIVNSTMKQV*

>G2065 (33..1124)

AACCACACAAAACAAAACAAAAAACATATTGATGGGGATGAAGAAGGTAAAGCTATCTT
TGATAGCTAATGAAAGATCAAGGAAAACATCCTTCATGAAGAGGAAAAACGGGATATTCA
AGAACTCCACGAGTTGTCAACTCTATGTGGTGTCCAAGCTTGTGCTCTCATCTATAGTC
CATTATACCGGTTCCAGAGTCATGGCCGTCAAGGGAAGGTGCTAAAAAGGTAGCTTCAA
AGTTTCTGGAGATGCCCGGACAGCCCGAACCAGGAAGATGATGGATCAAGAAACCCATC
TTATGGAGAGGATTACCAAAGCAAAAGAGCAACTAAAGAATTTGGCTGCTGAGAACCGAG
AATTACAGGTTAGACGATTATGTTTGTATTGTGTTGAAGGCAAAATGTCCAGTATCGTT
ATGATGCAAAAGACCTTCAAGATTTGCTATCTTGTATGAATCTATATCTCGATCAGCTTA
ACGGAAGGATCGAGTCCATTAAAGAAAACGGTGAGTCGTTGTTGTCTTCCGTCTCTCCTT
TTCCTACTAGAATTGGTGTGACGAAATGGTGATGAGTCGTTTTCCGACTCTCCTATTC
ATTCTACAACCTAGGGTTGTAGATACTCCTAATGCTACCAATCCTCATGTTCTTGC GGCG
ATATGACTCCTTTTCTTGATGCGGACGCAAATGCGGTAAGTCTCCAGTCGATTTTCTG
ATCATATTCAATATGAAAATATGAATATGAGTCAAAATCTGCATGAACCGTTTCAACACC
TTGTTCTCTACTAACGTTTGTGATTTTATCAAAATCAGAATATGAATCAGGTTCAATACC
AGGCTCCTAATAATCTGTTTAAATCAGATTCAACGAGAATTCTACAACATAAATTTGAATC
TGAATTTGAATCTGAATTCAAATCAGTATCTGAATCAACAACAATCATTATGAATCCGA
TGGTGGAACAACATATGAATCATGTTGGAGGGCGTGAAAGCATTCTTTCTGAGACAGAA
ACTACTACAACATACTAATCAACTACCAGCCGTTGATCTTGCTTCCACCAGTTACATGCCTT
CAACCACCGATGTTTATGATCCTTACATCAACAACAATCTCTAATCACAAAAGACGGAGA
TTTTCTAGTTTAA

>G2065 Amino Acid Sequence (domain in AA coordinates: TBD)

MGMKKVKLSLIANERSRKTSFMKRNKIGFKKLHELSTLCGVQACALIYSPFIPVPESWPS
REGAKKVASKFLEMPRTARTRKMDQETHLMERITKAKEQLKNLAAENRELQVRRFMFDC
VEGKMSQYRYDAKDLQDLLSCMNLVLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
DESFSDSPIHSTTRVVDTPNATNPHVLADMTPLDADANAVTAPSRFSDHIQYENMMS
QNLHEPFQHLVPTNVCDYFQONQNMNQVQYQAPNNLNFNQIQREFYNINLNLNLNSNQYL
NQQQSFMNPMVEQHMNHVGGRESIPFVDRNYNYNQLPAVDLASTSYMPSTTDVDPYIN
NNL*

>G2137 (77..1123)

GGGATTTGACTTTTAGCACTTCAAATCCAAAGCTAAAAGACAAAAAGAATAGAGGTTTCG
ATTTGCATCTCCATTAATGGGCATCGATCTTTCTCTTAAGCTCGAGGCCGAGGAGAAAA
GAAAGAGATAGAAGGATCGAAACATAGCCGTGAGAACAAAGAAGACGAAGAACATGATGC
TAGTGGTGAAGATGAAGTGAACAAATGGTGAAAGAAGACGAAGATGATCTTCTTTTAGG
TTTAAGAACCCGAGAAGAAGAAAACGAACGTGAAGAGCTCTTGCAGCTACAGATCCAGAT
GGAAAGTGTGAAAGAAGAGAATACTAGGTTGAGGAAGCTTGTGAGCAGACTCTTGAAGA
TTATCGTCATCTTGAGATGAAATTCCCGGTTATCGATAAAACCAAGAAGATGGATCTTGA
AATGTTCTTGGAGTACAAGGCAAACGATGTGTGGATATAACAAGTAAGGCTCGGAAAAG
AGGAGCTGAGAGATCTCCGTCAATGGAAAGAGAAATAGGGCTTTCACTTTCTCTAGAGAA
AAAACAGAAAACAAGAAGAGAGCAAAGAAGCTGTTCACTCTCATCACCAGATACAAATAG
TAGCAGCTTAGATATGAATATGCCACGTATCATTTTCATCTTCTCAAGGTAATAGAAAGGC
CAGGGTGTCCGTGAGGGCGAGATGTGAGACCGCAACAATGAATGATGGATGCCAATGGAG
GAAGTACGGTCAGAAAACCGGAAAGGGAATCCATGTCTCGAGCTTATTACCGATGCAC
CGTGGCTCCAGGATGTCCCGTTAGAAAACAGGTGCAAAGGTGTTTGAAGACATGTCAAT
ACTGATAACAACCTACGAAGGAACACATAACCATCCACTTCCGGTCCGAGCAACAGCCAT
GGCTTCCACTGCCTCTACTTCTCCATTCTTGTACTCGATTCCAGTGACAACCTCTCTCA
TCCTTCCATTATACCAAACCTCAAGCCATAGACTCTTCTTTGATTACATACCCACAAAA
TAGCAGCTACAACAATCGAACCATAGAAGCTTGAACCTTGTATGGTCCATCTAGAGGAGA
TCACGTTTCATCTTCTCAAACCGATTAAATTGGATGATGTAGAGTTTCTATATCTCTA
TGCTTGTCTTTGGTCCCATTTATTTGTCATTATGGATTCTTTGCCTTTCTTCTGTTCTC
GTTTCTAACATTTATGTTTCGTATA

>G2137 Amino Acid Sequence (conserved domain in AA coordinates:109-168)

MGIDLSLKLEAEEKKKEIEGSKHSRENKEDEEHDSASGEDEDEQMVKEDEDDSSSLGLRTRE
EENEREELLQLIQMESVKEENTRLRLKLEQTLQEDYRHEMKFPVIDKTKKMDLEMLGV
QGKRCVDITSKARKRGAERSPSMEREIGLSLSLEKKQKQESKEAVQSHHQRYNSSSLDM
NMPRIISSSQGNRKARVSVRARCETATMNDGCQWRKYGQKTAKGNPCPRAYYRCTVAPGC
PVRKQVQRCLEDMSSILITTYEGTHNHPLPVGATAMASTASTSPFLLLDSSDNLSPSYQ
TPQAIDSSLITYPQNSSYNRTIRSLNFDGPRGDHVSSSQNRLNWM*

>G746 (1..1311)

ATGGGTGAGGAGTTAGCTGACACAATGAACCTGGATTTGAATCTTGGGCTGGTCTGAG
TCTGATCTCCAACCTGCACCAAACGAGACTGTGAATTTGGCTGATTGGACTAATGACCCG
CCTGAGAGATCTTCTGAAGCTGTGACAAGGATCAGGACTCGGCATAGGACACGGTTCAGA
CAGCTTAATCTCCCGATCCCGGTTCTATCTGAAACCCATACCATGGCTATAGAGCTCAAC
CAGTTGATGGGAAATTCTGTAAATAGAGCTGCTATGCAGACTGGTGAGGGTAGTGAAAGA
GGCAATGAGGATTTGAAAATGTGTGAGAATGGCGATGGAGCCCTTGGGGACGGTGTATTG
GATAAGAAAAGCGGATGTCGAGAAAAGCAGTGGCAGCGACGGTAACCTTTTCGATTGTAAT
ATATGTTTGGATTTGTGGAAGGAGCCGGTCTCACCTGTTGTGGTCATCTTTACTGTTGG
CCTTGCTGTACCAATGGTTACAAATTTCCGATGCAAAGGAATGTCTGTTTGTAAAGGA
GAGGTGACCTCCAAAACCGTGACACCGATCTATGGACGTGGAACCAAGAGAGAAATT
GAAGAGAGTTTAGATACTAAGGTCCCCATGAGACCACACGCGAGACGATTGAGAGCTTG
AGGAATACAATTCAAAGGTCGCCTTTTACAATACCAATGGAAGAAATGATTAGACGTATA
CAGAATAGGTTTGACAGGGATTCAACCCAGTCCCTGATTTTAGTAACCGAGAGGCATCA
GAAAGAGTCAACGATCGAGCCAATTCGATCCTTAACCGGTTGATGACATCTAGGGGAGTT
AGATCAGAGCAGCAACGAGGTAGTGCTGCAGCAGCAGCATTGTGCGAGCATCAGAGGAT
ATTGATCTAAATCCAAACATATGCTCCTGATCTTGAAGGAGAAAGCAACACGAGATTCCAT
CCTCTCTTGATCAGGAGACAGTTACAGTCGCACCGAGTTGCAAGGATCTCGACTTCACT
TCTGCGTTGAGTTGAGCTGAGAGGCTTGTGGATGCGTATTTTAGGACTCATCCGTTGGGG
AGGAACCAACAAGAGCAAAACCATCATGCTCCTGTTGTGGTTGATGATAGAGACTCATTC
TCAAGCATTGCAGCTGTTTATAAATCTGAGAGTCAAGTGGATACTGCAGTTGAGATCGAT
TCTATGGCTCTTTTCGACATCGTCTCGAGGAGAAGGAATGAGAATGGTTTCGAGGGTTTCT
GATGTAGACAGTGCAGATTCTCGTCCGCCTAGGAGAAGGAGATTTACTTGA

>G746 Amino Acid Sequence (domain in AA coordinates: 139-178)

MGEELADTMNLDLNLGPGPESDLQPAPNETVNLADWINDPPERSSEAVTRIRTRHRTFR
QLNLPPIPVLSEHTMAIELNLQMLGNSVNRAAMQTGEGSERGNEDLKMCENGDGLDGV
DKKADVEKSSGSDGNFFDCNICLDLSKEPVLTCGHLIYCWPCLYQLQISDAKECPVCKG
EVTSTKVTPIYGRGNHKEIEESLDTKVPMPHARRIESLRNTIQRSPFTIPMEEMIRRI
QNRFDSDTPVPDFSNREASERVNDRANSILNRLMTSRGVRSEQNQASAAAAAIVAASED
IDLNPNIAPDLEGESNTRFHPILLIRRLQSHRVARISTFTSALSSAERLVDAFYRTHPLG
RNHQEQNHHPVVVDDRDSFSSIAAVINSESQVDTAVEIDSMALSTSSSRRRNENGSRV
DVDSADSRPPRRRRFT*

>G2701 (46..837)

GTGTTTGTAGTTGAAACTTATCTCTCCCTTTTTTTGTTTTTAGGTATGGAGACTCTGCAT
CCATTCTCTCACCTACCTATCTCTGACCACCGGTTGTTTCAAGAGATGGTGAGCTTA
CACAGCTCGAGTAGCGGTAGCTGACTAAAGAAGAGAAACAAGATGTTTCAACGAGCTCTT
GCGATATACGCTGAAGACTCGCCTGATCGCTGGTTTAAAGTTGCTTCCATGATCCCTGGA
AAGACTGTTTTTGATGTTATGAAGCAATATAGTAAGCTTGAAGAAGACGTTTTTCGATATT
GAAGCAGGACGTGTTCCCATTCCTGGTTATCCTGCAGCTTCTTCTCCCTTGGGGTTTGAC
ACGGACATGTGTGTAACCGGCCTAGTGGAGCTAGAGGATCTGATCAAGATCGAAAGAAA
GGAGTCCCTTGGACAGAGGAAGAACAACAGGAGATTCTGTTAGGCCCTTCAAGTACGGT
AAAGGAGATTGGAGAAACATATCGAGAAACTTCGTGGTGTCAAAGACGCCAACGCAAGTG
GCGAGCCACGCCAAAAGTATTACCAGAGACAGCTCTCCGGAGCCAAGGACAAAACGAGG
CCAAGTATCCATGACATCACAACCGGCAATCTTCTCAATGCCAATCTCAACCGTTCTTTT
TCCGATCATAGAGATATTCTCCCTGATTTAGGGTTTATCGATAAGGATGATACGGAGGAG
GGAGTAATATTTATGGGTGAGAAATCTCTCTTCAAGAAATCTGTTTTCTCCATCACCAACT
TCATTGGAAGCTGCCATTAACCTTCGCCGGAGAAAATGCTTTCAGTGCCGGAGCTTAAGGC
AACATAGAATCCCCAACTCAGCGGC

>G2701 Amino Acid Sequence (domain in AA coordinates: 33-81, 129-183)

METLHPPSHLPISDHRFVVQEMVSLHSSSSGSWTKEENKMFERALAIYAEDSPDRWFKVA

SMIPGKTVFDMVKQYSKLEEDVFDIEAGRVPIPGYPAASSPLGFDTDMCRKRPSGARGSD
QDRKKGVPTWTEEEHRRFLGLLLKYGKGWNRNISRNFVVSKTPTQVASHAQKYQRLSGA
KDKRRPSIHDITTNLLNANLNRSFSDHRDILPDLGFIDKDDTEEGVIFMGQNLSSENLF
SPSPTSFEAAINFAGENVFSAGA*

>G1819 (1..639)

ATGGAAGAGAACAACGGCAACAACCACTACCTGCCGCAACCATCGTCTTCCCAACTG
CCGCCGCCACCATTGTATTATCAATCAATGCCGTTGCCGTCATATTCAGTCCGCTGCCG
TACTCACCGCAGATGCGGAATTATTGGATTGCGCAGATGGGAAACGCAACTGATGTAAAG
CATCATGCGTTTCCACTAACCAGGATAAAGAAAATCATGAAGTCCAACCCGGAAGTGAAC
ATGGTCACTGCAGAGGCTCCGGTCTTATATCGAAGGCTGTGAGATGCTCATTCTTGAT
CTCACAATGCGATCGTGGCTTCATACCGTGGAGGGCGGTCCGCCAAACTCTCAAGAGATCC
GATACGCTCAGGATCCGATATCTCCGCCGCAACGACTCGTAGTTTCAAATTTACCTTC
CTTGGCGACGTTGTCCCAAGAGACCCTTCCGTCGTTACCGATGATCCCGTGCTACATCCG
GACGGTGAAGTACTTCTCCGGGAACGGTGATAGGATATCCGGTGTTTGATTGTAATGGT
GTGTACGCGTCACCGCCACAGATGCAGGAGTGGCCGGCGGTGCCTGGTGACGGAGAGGAG
GCAGCTGGGGAATTTGGAGGAAGCAGCGCGGTAATTGA

>G1819 Amino Acid Sequence (domain in AA coordinates: 46-188)
MEENNNGNNHYLPQPSSSQLPPPLYQSMPLPSYSLPLPYSPQMRNYWIAQMGNATDVK
HHAFPLTRIKKIMKSNPEVMNVTAEPVLISKACEMLILDLTMRSLHTVEGGRQTLKRS
DTLTRSDISAATRSFKFTFLGDVVRDPSVVTDDPVLHPDGEVLPPGTIVIGYPVFDNCN
VYASPPQMGEWPAVPGDGEEAAGEIGSSGGN*

>G1227 (372..1451)

TCTTCCGTGTGTTAACAGAAGTCCCCACAATTGTCTGTCTTCGCTGCGAGACAAAACCTGC
CACAGCCAATAATGTTTCTCTGAGGGACCTTGCTTCTGTGACAGACTCGCTCTCTCTCTC
TCTTCTTCTGCTCTGCTCAGCTCTCTCACCAACTCATCTTCAGTCCCTCAAACAAACATCTG
TTCTCATCTTTGTTTCTTTCTTTCTTTCTCATATCTCATTTTCAATTTTCCCAATTTCT
TCTTCAACATCTTCATAGCAATTTAAGACCACTATTCCATTATAAAGCTAACTGCTTTAG
AAACTCCTCACATTTATTTCTTCCCATCATTTGTTTGTAGAGAGGGAGAAAGAAAAAGAGC
TCAGCTTTCTGATGGAGAGGAGTATTCAAGGACAAAACAAGCTCTGTTGTTTGGACCAA
AAGTGAATGTGAGAAGAAGCCTACAAGTTCAAGAACTGTAGAGGATCATCAAAGCTTTG
CCCTTGAAGAGGAAGAACAACAACCTCTCAACTCCGAGCTTGCTGCAAGACACAACAATAC
CATTTCTACAAATGCTGCAACAAAGTGAAGACCCTTACCCTTTTGTTCATTCAAAGACC
CAAGCTTTCTAGCACTACTATCTCTCCAGACACTTGAAAAGCCTTGGGAACCTCGAAAAC
ACCTCCCACTAGAAAGTTCCAGAGTTTCAATCACCGATCCATTCTGAACCAACCACTACT
ATCATAATCCATCTTTGGAAGGAGTCAATGAAGCCATCTCAAACCAAGAACTTCCATTCA
ACCCACTAGAGAATGCGCGTTCAAGACGCAAGCGGAAAAACAACAACCTTGGCATCATTGA
TGACAAGAGAAAAGCGAAAGAGAAGAAGAACTAAACCAACAAGAACATAGAAGAGATAG
AGAGTCAAAGAATGACACACATTGCGGTTGAACGAAACCGCAGACGCCAAATGAACGTTT
ATCTGAACCTCACTCCGCTCCATCATTTCCATCTTTCATACATCCAGAGGGGAGACCAAGCGT
CAATAGTAGGAGGAGCAATAGACTTCGTAAAGATCCTAGAGCAACAGTTGCAATCCCTTG
AAGCACAAAAGAGAAGTCAACAGAGTGATGATAACAAAGAGCAAAATTCAGAAGATAACA
GTCTCAGGAACATTTGTCGAACAAGTTGCGTGCGAGTAATAAAGAAGAACAAAGTAGCA
AACTCAAATCGAAGCCACAGTGATAGAGAGTCACGTCAACCTAAAAATTCATGTACGA
GGAAACAAGGACAACCTTCTCAGATCAATCATATTGCTGGAGAACTTCGATTCACTGTTT
TTCATCTCAACATCACATCTCCGACCAATACATCTGTCTCTTATTCTTCAACCTCAAGA
TGGAAGATGAATGTAATTTGGGATCAGCGGATGAGATAACGGCGGCGATTCTGTCAGATTT
TCGACAGCTGATTGACTAATCCAAGTAAAAAGTAAAAATAAAAAAGAAACGTTTACTTTG
GTAACCTTCGTTTTCATGATTAAATCTTTATTTGGTTCGTATGTGATTGGAGTCTTCTCGG
CATGGAACCTTGACTTTGGTTTTAGGGTACTAGTCTCTACAGAAGCTGTGGTCCTTCTTTG
GATGC

>G1227 Amino Acid Sequence (domain in AA coordinates: 183-244)
MERSIQGQNKLCCLDQKVNVRSLQVQETVEDHQSFALLEEEQQLSTPSLLQDITIPFLQ
MLQSEDPSPLFSFKDPSFLALLSLQTLKPEWELNYLPHEVPEFHSPHSETNHYHNP
SLEGVNEAISNQELFPNPLENARSRRKRKNNNLASLMTREKRKRRTKPTKNIEEIESQR
MTHIAVERNRRRQMVHLNSLRSIIPSSYIQRGDQASIVGGAIDFVKILEQQLQSLEAQK
RSQQSDDNKEQIPEDNSLRNIISSNKLRSNKEEQSSKLKIEATVIESHVNLIKIQCTRKQG
QLLSIILLEKLRFVLHLNITSPNTSVSYFNLKMEDECNLGSADEITAAIRQIFDS*

>G2417 (118..1311)

CATACCGGTGGAAGATTCTGCTTTACTACGCTCTCCGCTTCTTCTTCTCCTCGATTTCGAT
TCTCCTCATGGGTTTATCATGAATTTTATAGGTTTGTAGTAATTCAGAACTCGAGTGATG
ATCCCGAATGATGATGATGATGCAAATCTATGAAGAATTATCCGTTAAATGATGATGAT
GCAAATCTATGAAGAATTATCCGTTAAATGATGATGATGCAAATCTATGGAGAATTAT
CCGTTAAGGTCAATTCCGACGGAGCTTTCACACACTTGTTTCATTGATACCACCTTCTTTA
CCAAACCCCTTCAGAAGCAGCAGCAGACATGTCCTTCAATTCAAGCTCAATCAAATCATG
GCAAGGCCTTGATATGCTCCCTGCCAATGGTGGAGCTGTTGGTCATAACCCCTTTTTTG
GAACCAGGATTCAACTGCCCCGAGACAACAGATTGGATTCCCTCTCCACTCCCCCATATT
TATTTTCTCTCGGGTTCTCCCAATCTAATAATGGAGGATGGTGTCTATTGATGAGATTCAC
AAACAAAGTGACTTGCCACTTTTGGTATGACGACTTGATTACCACTGATGAAGATCCACTC
ATGTCTAGTATCTTGGGCGATCTTCTCCTTGACACTAATTTCAACTCAGCTTCAAAGGTC
CAGCAACCAAGTATGCAATCGCAGATTCAACAACCCCAAGCTGTTCTGCAGCAGCCTTCT
TCTTGTGTGGAATTGCGCCCACTTGATAGGACAGTATCCTCAAACAGCAACAACAATAGC
AACAGTAATAATGCAGCAGCAGCAGCTAAGGGACGTATGCGTTGGACGCTGAACCTCAT
GAGGTTTTTGTGACGCTGTTAACCAGCTCGGTGGCAGTAATGAAGCAACTCCTAAAGGT
GTCCTGAAGCATATGAAAGTCGAAGGTTTGACTATTTTTTCATGTCAAAAGTCATTTGCAG
AAATATAGAACAGCTAAATATATACCAGTACCATCAGAAGGTTCCCGGAGGCAAGGTTG
ACACCGCTTGAGCAAATTACATCTGATGATACGAAACGTGGGATAGATATCACTGAGACT
CTGCGAATTTCAGATGGAACATCAGAAGAACTGCATGAGCAGCTTGAGAGTCTAAGAACA
ATGCAACTTCGGATAGAAGAGCAAGGAAAGGCGCTGTTGATGATGATTGAGAAGCAAAAT
ATGGGTTTCGGCGGACCAGAACAAAGGAGAGAAAAAAGTGCAGAAACGCTGAAAATGGT
TCAGAGGATCGGAATCCCCGCGGCCAAAGCGTCCGAGAAATGAAGAATGAAGGAAACCT
TTCTTCGGATGGTAGATCATAAACTGTGGTTTGGTGGAGTTGTAGAGTATGACTTATT
AGGAGTAGAGCTTTCAGTCTTCTTCAGGC

>G2417 Amino Acid Sequence (domain in AA coordinates: 235-285)

MI PNDDDDANSMKNYPLNDDDDANSMKNYPLNDDDDANSMENYPLRSIPTELSHTCSLIPPS
LPNPSEAAADMSFNSELNQIMARPCDMLPANGGAVGHNPFLPEPGFNCPETTDWIPSPPLPH
IYFPGSPNLIMEDGVIDEIHQSDLLPLWYDDLITDDEPLMSSILGDLILLDTNFNSASK
VQQPSMQSQIQPQAVLQPPSSCVELRPLDRTVSSNSNNNSNNNAAAAAKGRMRWTPPEL
HEVFVDAVNQLGGSNEATPKGV LKHMKEGLTIFHVKSHLQKYRTAKYIPVPSEGSPEAR
LTPLEQITSDDTKRGIDITETLRIQMEHQKLEHQLSRLTMQLRIEEQKALLMMIEKQ
NMGGFGPEQGEKTSAKTPENGSESESPRKRPRNEE*

>G2116 (104..1117)

TTCACTCCATCATTATCTCCATTGACATTGTTCTCAATTGCGAATAATAATCATAATTA
TTCACACAACCAAGCATTCATCTCTCAGATTCTCTTAAAAAATGGAGAAATCAGATCC
TCCACCAGTCCCAAAGCCCGCGCCACTATTATCCCCTCCTCCGATCCAATTCTTAATGC
CGATCCGATTCCATCTTCTTCTTCCACCGCCGATCTCGCTCCGACGATATGTCCATGTT
CATGTTTCATGATCCCTCTCCTCCGCGCACCACCTTCTTCCGACGACCTTCCCTCCGA
CGACGATCTCTTCTCTTCTTTCATCGATGTCGATAGCCTCACCTCTAATCCCAATCCCTT
TCAAAATCCTTCCCTCTCCTCCAACCTCCGTTTCCGCGCTGCTAATCCTCCTCCTCCTCC
TTCCTCTCGTCTCGCCACCGTCACAGCAATTCCGTTGACGCTGGATGCGCATGTATGC
CGGTGATATCATGGACGCTAAGAAAGCTATGCCTCCTGAAAACTCTCTGAGCTTTGGAA
CATCGATCCCAAACGCGCCAAAGGATTCTAGCGAATCGACAATCTGCAGCTCGATCCAA
AGAGAGAAAAGCTCGATACATTCAAGAACTTGAGCGCAAAGTTCAATCTCTTCAAACCGA
AGCTACCACTCTCTCTGCTCAGCTTACTCTCTACCAGAGAGACACAAATGGACTAGCAAA
CGAAAACACAGAGCTGAAACTTAGGTTGCAAGCAATGGAACAACAAGCTCAGCTTCGTAA
TGCTTTTAAACGAAGGTTTGAGGAAAGAGTTGAAAGGATGAAGATGGAGACAGGAGAAAT
CTCTGGTAATTTCAGATTCTGTTGATATGGGAATGCAGCAGATTTCAGTATTCTTCTCAAC
TTTCATGGCTATTCCACCATATCATGGCTCAATGAACCTCCATGATATGCAGATGCATTTC
TAGTTTCAATCCTATGGAGATGTCCAATTCTCAAAGCGTGTCCGACTTTCTACAGAACGG
CCGAATGCAAGGGCTGGAGATTAGTAGCAATAGCTCAAGCTTAGTCAAATCTGAAGGACC
TTCTCTCTCTGCTAGTGAGAGTAGCTCTGCCTATTGACGACAAGATTATGATGAGGCTCA
TTTTTCTG

>G2116 Amino Acid Sequence (conserved domain in AA coordinates: 150-210)

MEKSDPPVPKPGATIIPSSDPIPNADPIPSSSFHRRSRSDMSMFMDPLSSAAPPSS
DDLPSDDDLFSSFIDVDSLTSNPNPFQNPSSLSSNSVSGAANPPPPSSRPRHRSNSVDA

GCAMYAGDIMDAKKAMPPEKLSELWNIDPKRAKRILANRQSAARSKERKARYIQELERKV
QSLQTEATTLAQLTLYQRDTNGLANENTELKLRLQAMEQQAQLRNALNEALRKEVERMK
METGEISGNSDSFDMGMQIQYSSSTFMAIPPYHGSMNLHDMQMHSSFNPMEMSNSQSVS
DFLQNGRMQGLEISSNSSSLVKSEGPSLSASESSAY*

>G647 (1..948)

ATGATGATCGGCGAAAATAAAAACCGGCCACATCCAACGATCCATATCCCTCAATGGGAT
CAAATCAACGATCCAACGGCCACAATCTCTTACCATTCTCTCCGTCAACCTTAACAGC
GTTAACGACTACCCACACTCTCCGTACCGTATCTCGACTCCTTCGCTTCTCTCTCCGT
TACCTCCCGTCAAACGAGTTAAACAAACGATTGAGACTCATCAAGTGGCGACGAGTCATCA
CCACTCACCGACTATTCTCTCCGACGAGTTTCGCATCTACGAGTTCAAAATCCGCGCA
TGCCTCGAGGTCGATCTCATGATTGGACGGAGTGTCCGTTTCGCACATCCCGGAGAAAAA
GCTCGACGACGTGATCCGAGAAAGTTTTCATTACTCCGGCACCCTTGTCTGAGTTTCTGT
AAAGGAAGTTGTAGAAGAGGTGATTCTGTGTGAGTTCTCTCATGGAGTTTTCGAGTGTGG
CTCCATCCTTCTCGTTACCGTACTCAGCCGTGTAAAGACGGAAGTAGCTGCCGAGAGAAGA
ATCTGTTTCTTCGCTCATACGACGGAGCAGTTACGTGTATTACCTTGTTCGTTAGATCCA
GATCTTGGATTCTTCTCAGGATTAGCTACTTCTCCGACTTCGATTCTTGTTCCTCTCTCG
TTTTCAACACCGTCGGAATCTCCGCCGCTTCTCCGAGTACCGTGAACCTTATTGCGTCG
ATGAGGAAAATGCAATTGAACGGAGGTGGTTGTTTCGTGGAGTTCTCCGATGAGATCTGCA
GTTAGGTTACCTTTTTCTGTCGTCTCTGCGTCCGATTACGGCGCAACGTGGCCGAGGATA
AGAGAGTTTGAATCGAAGAAGCTCCGGCGATGGAATTTGTGGAATCTGGGAAAGAGCTG
AGAGCGGAGATGTATGCAAGACTCAGTAGAGAGAACTCACTCGGTTGA

>G647 Amino Acid Sequence (domain in aa coordinates: 77-192)
MMIGENKNRPHPTIHIPQWDQINDPTATISSPFSSVNLNSVNDYPHSPSPYLD SFASLFR
YLP SNELTNDSDSSSGDESSPLTDSFSSDEFRIYEFKIRRCARGRSHDWTECPFAHPGEK
ARRRDP RK FHYSGTACPEFRKGSCRRGDSCEFSHG VFECWLHPSRYRTQPKDGTSCRRR
ICFFAHTTEQLRVLPCSLDPLDGLFFSGLATSPTSILVSPSFSPSESPPLSPSTGELIAS
MRKMQLNGGGCSWSSPMRSVRLPFSSSLRPIQAATWPRIREFEIEEAPAMEFVESGKEL
RAEMYARLSRENSLG*

>G974 (377..1162)

AAAAAAAAAGTTGATATACTTTCTGGTTTTCTCCTTAACTTTTATTCTTTACAAATCCAT
CCCCCTTAGATCTGTTTATTTCCCGCTACTTTGATTCAATTTCTGTTAGTAATCTGTCTTT
CGTATAGAAGAAAATGATTCTTGGTTTGTATTTTCTTAAAGAGATCAATCTTTTTTTTA
TTTTTGATCTTCTGTGTTTTTTTCTTTTGTAGAAATTAATCGTTTGTGAGGGTATTTTTT
TTAATTCCTCCTCTCAGAAATCTACACAGAGGTTTTTTTATTTTATAAACCTCTTTTTCTG
ATTTTCTTGAAAACAAAAATCCTGTTCTTTACTTTTTTTACAAGAACAAGGAAAAAAA
TTTCTTTTTATTAGAAATGACAACTTCTATGGATTTTTTACAGTAACAAAACGTTTCAACA
ATCTGATCCATTCCGTGGTGAATTAATGGAAGCGCTTTTACCTTTTATCAAAAGCCCTTC
CAACGATTCATCCGCGTTTTCGCTTCTCTACCCGCTCCAATTTATACAGGGTCGGATCT
CCACTCATTTTCTCACCATCTTAGTCCTAAACCGGTCTCAATGAAACAAACCGTACTTC
CGCGGCTAAACCGACGAAGCTATACAGAGGAGTGAGACAACGTCACTGGGGAAAATGGGT
GGAGGAAGCTGCTTTAGCTTATGACAAGCGCGGTATAAGCTCCGAGGAGATTTTTCGCG
GCTTAATTTCCCTGATCTCCGTCTAAGCAGAGTATCAACCTCTTCAATCATCAGTCGA
CGCTAAGCTTGAAGCTATTTGTCAAACTTAGCTGAGACGACGAGAAACAGGTGAGATC
AACGAAGAAGTCTTCTCTCGGAAACGTTTCAACCGTCGAGTGAAACTACCGGAGGA
GGACTACTCTAGCGCCGGATCTTCGCCGCTGTTAACGGAGAGTTATGGATCTGGTGGATC
TTCTTCGCCGTTGTCGGAGCTGACGTTTGGTGATACGGAGGAGGAGATTAGCCGCCGTG
GAACGAGAACGCGTGGGAGAAGTATCCGTCGTACGAGATCGATTGGGATTCGATTCTTCA
GTGTTCCAGTCTTGTAATTAGATGTTGCCATAGGGGTATTTTAGGGACTTTAGAGCTCT
CTGCGATGGAGTTTTTGGTCATTGCAGAGATTTTATTATTATTAAAGGGGGTTTGTATGT
TAATATCAAATAAGTTTATCTACTTTGATGTTAATTAGTGTTAATCTCTGCGTCGGTCCA
AGCTGTTTTTTTTTGGCATGCTTCGACCGTGTGAGATTTCTTATGTAATTTTTGTAGTTC
CTTGATTTTCTTAGTTCAAGTTAAATTGGCACAAAAA

>G974 Amino Acid Sequence (domain in AA coordinates: 81-140)
MTTSMDFYSNKTFFQSDPFGGELMEALLPFIKSPSNDSSAFASFSLPAPISYGSDLHSFSH
HLSPKPVSMKQTGTSAAKPTKLYRGVRQRHWGKWVAEIRLPRNRTRLWLGTDFDTAEAAAL
AYDKAAYKLGRDFARLNFPLRHNDEYQPLQSSVDKLEAICQNLAEETTQKQVRSTTKSS

SRKRSSTVAVKLPEEDYSSAGSSPLLTESYSGSGSSSPLSELTFGDTEEBIQPPWENAL
EKYPSYEIDWDSILQCSSLVN*

>G1419 (27..692)

GAAGACTCCAACATAATTCATCATCTATGGCTTCTTCACATCAACAACAGCAAGAACAAG
ACCAGTCAGCTTTAGATCTCATAACCCAAACACCTTCTTACTGATTTCCCTTCCTTAGACA
CCTTTGCCTCCACCATCCACCCTGCACCACCTCAACTCTAAGCCAACGCAAACCACCTC
TTGCCACTATAGCAGTTCTTACTACTGCACCGGTGGTTCAAGAGAATGATCAAAGGCATT
ACAGAGGCGTCAGGAGAAGACCATGGGGTAAGTATGCGGCTGAGATCAGAGACCCAAACA
AGAAAGGTGTTTCGTGTCTGGTTAGGCACCTTTTGACACAGCCATGGAAGCTGCAAGAGGTT
ATGACAAGGCAGCTTTTAAACTACGAGGAAGCAAAGCTATTCTTAACTTCCCACTTGAAG
CAGGAAAGCATGAGGACTTGGGAGACAACAAGAAGACTATTTCTTTAAAGCAAAGAGGA
AGAGACAGGTGACGGAGGATGAAAGCCAGCTGATCAGCCGTAAAGCTGTTAAGAGGGGAG
AAGCTCAGGTTTCAAGCTGATGCTTGTCCATTAACGCCATCAAGTTGGAAGGGGTTTGGG
ACGGAGCAGACAGTAAAGACATGGGAATATTTCCGTGCCTCTGTTATCTCCTTGTCCAT
CTCTTGGACACTCTCAACTCGTAGTTACTTAAGCTTCAGAGGGTCAAACCTGGAAAAATC
AACATTGGATTGTTTTCAAAGCTTCTAGATTAGCTGATTGTAAAAAATGTTTTACTATA
TTCATTCACTTCTTCTTAAATGCAATTCCTTCTACCTTCC

>G1419 Amino Acid Sequence (domain in AA coordinates: 69-137)

MASSHQQQQEQDQSDLDLITQHLLTDFPSLDTFSTIHHCTTSTLSQRKPPLATIAVPTT
APVVQENDQRHYRGVRRRPWGKYAAEIRDPNKKGVVRVWLGTFDTAMEAARGYDKAAFRLR
GSKAILNFPLEAGKHEDLDGNKKTISLKAKRKRQVTEDESQILSRKAVKREEAQVQADAC
PLTPSSWKGFWDGADSKDMGIFSVPLLSPCPSLGHSLQLVVT*

>G1634 (22..855)

TTATCTCGTAGCCTTTTAAACGATGGAGACTCTGCATCCACTACTCTCGCACGTGCCAACT
TCTGACCACCGGTTTGTAGTTCAAGAGATGATGTGCTTGCAAAGCTCGAGCTGGACTAAA
GAAGAGAACAAGAAGTTTGAGCGAGCTCTTGCTGTCTACGCTGATGACACGCCTGATCGC
TGGTTCAAAGTTGCTGTATGATCCCTGGAAGACCATATCAGATGTCATGAGGCAATAC
TCTAAGCTTGAAGAAGACCTCTTCGATATCGAAGCAGGACTTGTCCCGATCCCGGGTTAC
CGTTCAGTTACTCCTTGTGGATTGATCAGGTTGTGAGTCCACGTGACTTTGATGCGTAT
CGTAAACTTCTAATGGAGCCAGAGGATTGATCAAGACGTAGGAAAGGAGTTCCATGG
ACGGAGGAAGAACAACAGGAGATTCTTGTAGGGCTTCTCAAGTATGGGAAAGGAGATTGG
AGAAAACATATCGAGGAACCTTGTGGGATCAAAAAACCAACTCAGGTTGCAAGTCATGCC
CAAAAGTACTACCAAAGACAGCTTTCCGGTGCGAAAGACAAACGACGGCCTAGCATTCAC
GACATCACACCGTCAATCTTCTCAATGCCAATCTTAGCCGTCCATCGTCTGATCACGGT
TGCTTAGTCTCAAAACAGGCCGAGCCGAAACTAGGGTTACCGACAGGGATAATGCAGAG
GAGGGAGTTATGTTTCTTGGTCAGAATCTATCCTCGGTCTTCTTCTTCTACGATCCTGCC
ATTAAGTTTTCCGGAGCAAATGTTTACGGTGAAGGAGGTTACTGTATCTCACAAGATCTT
GAAACGAGAAATGAGAATTTTGAATTTTAACTATTGCAACGAAACCATAATTGC

>G1634 Amino Acid Sequence (domain in AA coordinates: 129-180)

METLHPLLSHVPTSDHRFVVQEMMCLQSSSWTKEENKKFERALAVYADDTDPDRWFKVAAM
IPGKTISDVMRQYSKLEEDLFDIEAGLVPIPGYRSVTPCGFDQVVSPRDFDAYRKLPGA
RGFDQDRRKGVPTWTEEHRRFLLGLLKYKGDWRNISRNFVGSKTPTQVASHAQKYYQRQ
LSGAKDKRRPSIHDITTVNLLNANLSRPSSDHGCLVSKQAEPLGFTDRDNAEEGVMFLG
QNLSSVFSSYDPAIKFSGANVYEGGGYCISQDLETRK*

>G1637 (1..954)

ATGGTGAAGGAGACGGTGACGGTGGCGAAAAACGTGCTCACACTGTGGCCATAATGGCCAT
AACGCACGGACTTGTCTCAACGGCGTTAATAAGGCAAGTGTTAACTGTTTCGGCGTTAAT
ATATCGTCTGATCCGATTAGGCCCGCTGAGGTAACGGCGTTAAGGAAGAGTCTTAGTTG
GGAAACCTTGATGCTCTTCTCGCTAACGATGAAAGTAACGGTAGCGGTGATCCTATCGCC
GCCGTTGATGATACCGGTTATCATTCCGATGGTCAGATTCAATCCAAGAAGGGTAAACT
GCTCATGAGAAGAAAAAGGGGAAGCCATGGACGGAAGAAGAACATCGTAATTTCTTAATC
GGTTTAAACAACTCGGAAAAGGAGATTGGAGAGGCATTGCAAAGAGTTTCGTGTGACACA
AGAACACCAACACAAGTCGCAAGTCATGCTCAGAAATATTTTATTAGGTTAAACGTTAAC
GACAAGAGAAAAAGACGTGCTAGTCTCTTGTACATCTCTCTCGAAGATCAGAAGGAGAAA
GAGAGGAAGCTCTCAAGATGCTTCAACAAAGACTCCACCTAAACAACCAATAACCGGAATT
CAACAACCGGTAGTACAGGCTCATACTCAAACCGAGATTTCGAACAGGTTTCAGAATTTA
TCAATGGAGTATATGCCAATCTACCAACCCATACCACTTACTACAACCTTTCCACCTATT

ATGTACCATCCAAATTATCCAATGTACTATGCCAACCTCAAGTACCGGTTAGGTTTGT
CATCCTTCTGGTATACCTGTTCCAAGACATATACCGATTGGTTTGCCTCTGTCTCAACCG
AGTGAAGCTTCTAATATGACAAATAAGACGGTTTGGATCTTCATATCGGTTTGCCTCCA
CAAGCTACTGGAGCTTCTGACTTGACTGGTCATGGCGTTATTCATGTGAAATGA
>G1637 Amino Acid Sequence (domain in AA coordinates: 109-173)
MVKETVTVAKTCSHCGHNGHNARTCLNGVNKASVKLFGVNISSDPIRPEVTALRKSLSL
GNLDALLANDESNGSGDPIAAVDDTGYHSDGQIHSKKGKTAHEKKKGKPWTEEEHRNFLI
GLNKLKGKDWRIAKSFVSTRPTQVASHAQKYFIRLNVNDKRKRRLASLFDISLEDQKEK
ERN SQDASTKTPPKQPIITGIQPVVQGHQTQTEISNRFQNL S MEYMPIYQPIPPYNFPPPI
MYHPNYPMYANPQVPVRFVHPSGIPVPRHIPIGLPLSQPSEASNMTNKDGLDLHIGLPP
QATGASDLTGHGVIHVK*
>G1818 (601..1161)
TAACAAATCAAATAATTAGAGAAATAACCAAAATTTAACTTTTAGAGGGACTACAGGATT
TGTACTTTGTACATTTCATATATTATTGTTATATATCGTTTCATACATTAATTTGAACCAA
TGTAATAAAGTAAATTTCAATTTAACATCATGAGCAAATTCCTATTAAATTTCTCTTAA
AATTTTGAGCAAATTATGCTTTCACATTTTAACATTTGAAAACATCATTTTAAACAAGATA
TTCAAACCTAAGTTTGTACAGCAAATTTTAACCTTCAATTTTATAGAGAAAAAGGTAT
TTTTTTTTTTGTTTCATTTTATAAGACTATTATTGTTATATAATATACACTTTAAGTA
AAAACAAATCTCTTTCTTTTCTTCTTATAATACCAACCACAAGTCTGTCAGTCACACA
CATACAGTTAATAACATTTAAATATTCTTAACAACTACTAAATAGGTTGAGATTTCATATA
TGTAAGAGATCACTTCTTAATCTTATCCTACCATATCTTATATACGCTTAATTTTCCTT
TATATATGCAAACCTCCACATAAAAAATATCTCAAACCCAAACACTTCAAACAAAAA
ATGGAGAACAAACAACAACCAACCAACAGCCACCGAAAGATAACGAGCACTAAAGAGT
TTCTGGTCAAAGGGGATGGAAGGTGACTTGAATGTCAAGAATCACGAGTTCCCCATCTCT
CGTATCAAGAGGATAATGAAGTTTGATCCGGATGTGAGTATGATCGCTGCTGAGGCTCCA
AATCTCTTATCTAAGGCTTGTGAAATGTTTGTGTCATGGACCTCACGATGCGTTCATGGCTC
CATGCTCAAGAGAGCAACCGACTCACGATACGGAAATCTGATGTTGATGCCGTAGTGTCT
CAAACCGTCATCTTTGATTTCTTGGCTGATGATGTCCCTAAGGACGAGGGAGAGCCCGTT
GTCCGCGCTGCTGATCCTGTGGACGATGTTGCTGATCATGTGGCTGTGCCAGATCTTAAC
AATGAAGAACTGCCGCCGGGAACGGTGATAGGAACCTCCGGTTTGTACGGTTTAGGAATA
CACGCGCCACACCCGCAGATGCCTGGAGCTTGGACCGAGGAGGATGCGACTGGGGCAAAT
GGAGGAAACCGGTGGGAATTAATATTGGATTGGGTTTGTAAACCGCTGTTGTGAGAATTT
GAATTTCTTTTGGATTCTGCTTATGTTTTCATGTTTATGTTTTTTAGTTGTTGAATGTA
TTTCTGTTGTTTTGTCCAAAAAAGAAATGTATTCTGTTGTTGTCTTTCAAATGA
ATCTAATGGTTTATGAATATTGGCTTTAGATTAATTTATGCATACAAAACACAAGGATT
ACGGATAAAAAAGTCTCAGTTTACCCATGGAAACATAATCTTCTAGTGATTCTTTATGA
GAGTAGAAAAGAAATCATATATTATAATCTATTTTCATAAGAGATAGGGTACTGTAAACAAG
GATGTTTATTTCGGCTATTTCTTTTTTTTTTAATCACTTTTACTTGTCAAGACTCTTTTGT
GTTTGCAGCTTTTGTGATTACATTCTAGAGGCAACAAGATCCAGAGATCTAGCAAAA
AAAACCTATTTTGAAACCTGAATCTATTTAAAAATTTTCCAACCTCATTTTTCGTTCTTA
TTCTTTGTTTTCACCGGAATTTGGCGCACAAACGATTATTGTAATTTTGTCTTTCAAG
>G1818 Amino Acid Sequence (domain in AA coordinates: 36-113)
MENNNNNHQPPKDNEQLKSFWSKMEGDLNVKNHEFPISRIKRIMKFDPDVSMIAAEAP
NLLSKACEMFVMDLTMRSWLHAQESNRLTIRKSDVDAVVSQTVIFDFLRDDVPKDEGEVP
VAAADPVDDVADHVAVPDLNNEELPPGTVIGTPVCYGLIHAPHQMPGAWTEEDATGAN
GGNGGN*
>G1820 (1..609)
ATGGCTGAGAACAAACAACGCGGACAAACATGAACAACGACAACCAGCAACCA
CCGTCGTACTCGCAGCTGCCGCGATGGCATCATCAACCCCTCAGTTACGTAATTACTGG
ATTGAGCAGATGGAACCGTCTCGGATTTCAAAAACCGTCAGCTTCCATTGGCTCGAATT
AAGAAGATCATGAAGGCTGATCCAGATGTGCACATGGTCTCCGCAGAGGCTCCGATCATC
TTCGCAAAGGCTTGCGAAATGTTTCATCGTTGATCTCACGATGCGGTCTGGCTCAAAGCC
GAGGAGAACAAACGCCACACGCTTCAGAAATCGGATATCTCAACGCAGTGGCTAGCTCT
TTCACCTACGATTTCTTCTTGATGTTGTCCCTAAGGACGAGTCTATCGCCACCGCTGAT
CCTGGCTTTGTGGCTATGCCACATCCTGACGGTGGAGGAGTACCGCAATATTATTATCCA
CCGGGAGTGGTGATGGGAACTCCTATGGTTGGTAGTGGAATGTACGCGCCATCGCAGGCG
TGCGCCAGCAGCGCTGGTGACGGGGAGGATGATGCTGAGGATAATGGAGGAAACGCGCGC

GGAAATTGA

>G1820 Amino Acid Sequence (domain in AA coordinates: 70-133)
MAENNNNNNGDNMNNNDNHQQPPSYSQLPPMASSNPQLRNYWIEQMETVSDFKNRQLPLARI
KKIMKADDPDVHMVMSAEAPIIFAKACEMFIVDLTMRSLKAEENKRHTLQKSDISNAVASS
FTYDFLLDVPVKDESIATADPGFVAMPHPDGGGVPQYYYYPPGVVMGTPMVGSGMYAPSQA
WPAAAGDGEDDAADNNGGGGNN*

>G1903 (1..1200)

ATGTCTAAATCTAGAGATACGGAGATAAAGTTGTTTGGGAGGACAATCACATCTCTTTT
GATGTGAATTGTTATGATCCGTCGTCGTTGTCCCTGTTACAGATGTTTCTCTGATCCA
AGCAAGGAGGATTTCGTCTTCTTCTTCATCTTCTTGTCTCCAACATATTGGACCAATCAGG
GTTCCGGTTAAAAAAGTGAGCAAGAGAGTAACAAATTCAAAGATCCATATATATTATCC
GATCTAAACGAACCACCAAAAGCAGTATCTGAGATTTCATCACCAAGAAGTTCCAAGAAC
AACTGTGATCAACAGAGCGAGATCACAACAACAACCTACCACAAGTACTACATCAGGAGAG
AAATCAACGGCTCTCAAGAAACCGGACAAGCTTATCCATGTCTTAGATGTGAAAGCGCA
AACACCAAAATTCTGTTATTATCAACAACATAACAAGTGAACCAGCCACGTTACTTCTGCAGG
AACTGTGACAGGATTGGACAGCTGGTGGATTATGAGGAACGTTCTCTGTTGGCTCAGGT
CGTCGCAAGAACAAGGATGGCCTTCTTCAAACCTACTTGCAGTCACTTCTGAGGAT
TGTGATAATAATAACTCGGGGACGATCCTTAGTTTCGGTTCTTCCGAGTCTTTCGGTTACA
GAGACTGGTAAGCATCAGTCAGGTGATACAGCAAAGATAAGTGCTGATTACGTTTCTCAA
GAAAATAAAAGCTACCAAGGGTTTCTTCTCCGCAAGTAATGTTACCTAATAATTCTTCT
CCTTGGCCTTACCAATGGAGTCCAACGGGTCTTAACGCTAGTTTCTACCCTGTCCCTTC
TACTGGGGATGCACGGTTCCGATATACCCTACCTCAGAGACTTCATCATGTTTAGGAAAA
CGGTCAAGAGATCAAACTGAAGGAAGAATCAATGATACTAATACACAATAACTACTACA
AGAGCAAGATTGGTCTCAGAATCTCTTAGAATGAATATCGAAGCTAGTAAGAGCGCTGTG
TGGTCTAAGTTACCGACAACAAACCCGAGAAAAACCGCAAGGATTACGTTTGTCTCAATGGA
TTTGACACAAGGGAAAAACAGCAACAGAAAGTAGCTTGGTCTCCGAACTTCTCACAGTCTA
CAAGCAAACCTTCGACGCGATGTCTAGAGCTATGAACCTTCAGGAGAGACTTCACAACTA

>G1903 Amino Acid Sequence (domain in AA coordinates: 134-180)

MSKSRDTEIKLFGRTITSLLDVNCYDPSSLSPVHDVSSDPKEDSSSSSSSCPTGPIR
VPVKKSEQESNKFKDPYILSDLNEPPKAVSEISSPRSSKNNCDQOSEITTTTTTSTTSGE
KSTALKKPKDLIPCPRCESANTKFCYNNYNVNQPRYFCRNCQRYWTAGGSMRNPVPGSG
RRKNKGWPSNHYLQVTSEDCDNNSGTTLSFGSSESVTETGKHQSGDTAKISADSVSQ
ENKSYQGFLPNQVMLPNNSPWPYQWSPTGPNASFYVPFVYGCTVPIYPTSETSSCLGK
RSRDQTEGRINDTNTTTTTRRLVYESLRMNIASKSAVWSKLTPKEKTKQGFSLFNG
FDTKGNSNRSSLVSETSHSLQANPAAMSRAMNFRESMQO*

>G371 (1..582)

ATGGAGATTGAGAAGGATGAGGACGACACAAACATTGGTTGATTCTGGAGGAGACTTCGAC
TGCAACATATGTTTGGATCAGGTTTCGAGACCCGGTCTGTGACTTTATGTGGCCACCTGTTTT
TGTTGGCCCTGCATTCAAAAGTGGACTTATGCGTCCAACAATTCAAGACAACGAGTTCGAT
CAATACGATCATAGAGGGGAACCACCAAATGTCCGGTATGCAAACTCTGATGTCCTCCGAG
GCTACGCTTGTCCCGATCTACGGACGAGGACAGAAAGCTCCCAGTCCGGTTCAAATGTA
CCGAGCAGACCAACTGGTCCGGTTTTATGACTTAAGAGGAGTTGGTCAACGTTTAGGAGAA
GGGAGAGTCAACGTTTACATGTATAGAAATGCCTGATCCGGTGATGGGTGTGGTATGCGAA
ATGGTATACCGGAGACTATTTGGAGAGTCTTCCGACAAATGGCACCTTACC GCGATATG
AATGTCCGGTCTAGGCGACGGGCAATGCAGGCTGAGGAGTCAATTAAGCAGAGTCTACTTG
TTTCTACTTTGCTTCATGTTTATGTGTCTATTTCTCTTCTAA

>G371 Amino Acid Sequence (domain in aa coordinates: 21-74)

MEIEKDEDDTTTLVDSGGDFDCNICLDQVRDPVVTLCGHLFCWPCITHKWTYASNNRQRVD
QYDHKREPPKCPVCKSDVSEATLVPIYGRGQKAPQSGSNVPSRPTGPVYDLRGVGQRLGE
GESQRYMYRMPDPVMGVVCEMVYRRLFGESSNMAPYRDMNVRSRRRAMQAEESLSRVYL
FLLCFMFMCLFLF*

>G597 (255..1310)

AAAATTCTCCTGTAATAATTAAATATTATAAAAGTGGTTTCTTTTCATTATATGTTTATAT
AATTTTCATCTTTAATCTTAAATCTGTTAACCTTAATGCGCGATCCGCTTTTCTAAAGT
TTTGTGAGAGAGAAGAGATCTAAAAAATCCACAATTTTGTCAAATCTTGGAGTTAAAT
GCTGAATTTTAGGCCCTTGTTGCTTAGATTTATGGCTTAAAGTTTCAAACTTTTCATTGGA
TATGTGAGAAGAAATGTCAGGATCTGAGACGGGTTTAATGGCGGCGACCAGAGAATCAA

TGCAATTTACAATGGCTCTCCACCAGCAGCAGCAACACAGTCAAGCTCAACCTCAGCAGT
 CTCAGAACAGGCCATTGTCTATTCGGTGGAGACGACGGAAGTCTCTTTACAAGCAGCCGA
 TGAGATCAGTATCACCACCGCAGCAGTACCAACCCAACTCAGCTGGTGAGAATTCTGTCT
 TGAACATGAACCTTGCCCGGAGGTGAGTCTGGAGGCATGACTGGAAGTGGAGTGGAGCCAG
 TGAAAAAGAGGAGAGGTAGACCGAGGAAATATGGGCCTGATAGTGGTGAATGTCACTTG
 GTTTGAATCCTGGAGCTCCTTCTTCTCACTGTCAGCCAACTAGTAGCGGCGCGATGGAG
 GAGAGAAGAAGAGAGGAAGACCTCTGGTTCTTCTAGCAAAAGGCTCAAGCTTCAAGCTT
 TAGGCTCGACTGGAATCGGATTACGCCTCATGTACTTACCGTGCTGGCTGGAGAGGATG
 TATCATCCAAGATAATGGCGTTAACTCATAATGGACCCCGTGCTGTGTGTGTCTTGTCTG
 CAAATGGAGCCATCTCCAATGTGACTCTCCGCCAGTCTGCCACATCCGGTGGAACTGTTA
 CATATGAGGGGAGATTTGAGATTCTGTCTTTATCGGGATCTTTCCATTTGTGAGAGAACA
 ATGGTCAAAGAAGCAGGACGGGAGGTCTAAGCGTGTCATATCAAGTCCGGATGGTAATG
 TCCTCGGTGGCAGTGTAGCTGGTCTTCTTATAGCAGCATCACCTGTTTCAAGTTGTGTG
 GGAGTTTCTTACCAGACGGAGAAAAAGAACCAAAACAGCATGTGGGACAAATGGGACTGT
 CGTCACCCGTATTACCGCGTGTGGCCCCAACGCAGGTGCTGATGACTCCAAGTAGCCAC
 AATCTCGAGGCACAATGAGTGAGTCATCTTGTGGAGGAGACATGGAAGCCCTATTTCATC
 AGAGCACTGGAGACCTTACAATAACACCATTAACATGCCCTGGAAGTAGCCAAGTGATC
 TGTGTCCGCTTAAAAACCAACAACCTCCCGTTATTAGAGTGATTTATTTCTACATTGGTT
 TAGACTTTCTAGTTCTGATGGTTATTTCTACAGTTGGTTTAGACTTTCTAGTTCTGTTCA
 GACAAAAGGAGTTTGATAAATTGACCGACCTATTTTGTGTGTTGAGGTACTTTCAGAAC
 CATAGGTGTTTCAGAAATTAGAATGTTCTGTTTAAAAAA
 >G597 Amino Acid Sequence (domain in AA coordinates: 97-104,137-144)
 MSGSETGLMAATRESMQFTMALHQQQHSQAQPPQSQNRPLSFGDDGTALYKQPMRSVS
 PPQQYQPNASGENSVLNMNLPGGESGGMGTGTGSEPVKKRRGRPRKYGPDSGEMSLGLNPG
 APSFTVSPSSGGDGEKKRRRPPGSSSKRLKLQALGSTGIGFTPHVLTVLAGEDVSSKI
 MALTHNGPRAVCLVLSANGAISNVTLRQSATSGGTVTYEGRFEILSLSGSFHLLNNQORS
 RTGGLSVSLSSPDGNVLGGSVAGLLIAASPVQIVVGSFLPDGEKEPKQHVGMGLSSPVL
 PRVAPTQVLMTPSSPQSRGTMSESSCGGHHGSPIHQSTGGPYNNTINMPWK*
 >G1009 (28..1704)
 AAAAAAAAAAAAAAACCATTCCCAAAGATGAAGAACAATAACAACAAATCTTCTTCTTCT
 TCTAGCTATGATTCTTCTTTGTCTCCTTCTTCTTCTCATCTCCTCCCACCAGAACTGGCTC
 TCTTTCTCTCTCTCCAACAATAACAACAACCTTCAATTCTTCTCAAACCTAATCTCACT
 TCCTCCACATCAGATCATCATCTCACCCTTCTCACCCTCTCTCTTTCAAGCTTTC
 TCCACTTCTCCAGTCCGACGCAAGATGGGTACCCGGGAGTTTACCCAGCGATGCCACG
 GCGGTTCTTTCCGTATACCCCGCGGCTCTAAACTTGAGAACTTCTCGGCGGAGGAGCC
 TCAACGACGACAACAAGACCAATGCAACAAGTGCAATCTCTTGGCGGCGTTGTCTTCTCT
 TCCGACCTACAGCCACCGCTTCATCTCCGTCCGCGCGGAGATCTACGACTCTGAGCTC
 AAGTCAATAGCCGCTAGCTTCTTAGGAACTACTCCGGTGGACACTCGTCCGAGGTCTCT
 AGCGTACATAAACAACAACCGAATCTCTAGCTGTCTCAGAGGCTTCGCCTACTCCGAAG
 AAGAACGTAGAGAGTTTGGACAACGTACCTCGATTTATAGAGGAGTCACAAGACATAGA
 TGGACTGGAAGATACGAAGCTCATCTATGGGATAATAGTTGCCGAAGAGAAGGCCAAAGC
 AGAAAAGGAAGACAAGTTTATTTAGGTGGTTATGATAAGGAAGATAAAGCAGCTAGAGCT
 TACGACCTTGACGCTCTTAAGTATTGGGGTCTTACAACACTACGACTAATTTCCCGATATCA
 AATTACGAATCTGAACCTGAAGAAATGAAACACATGACTCGACAAGAGTTCGTTGCTTCT
 TTAAGACGGAAGAGCAGTGGATTCTCTAGGGGTGCCTCCATGTACAGAGGCGTCACTAGA
 CATCATCAGCATGGTCGATGGCAGGCACGAATTGGAAGAGTTGCAGGCAACAAGACCTT
 TATCTTGGCACATTTAGCACTCAAGAGGAAGCTGCAGAAGCTTATGATATAGCAGCGATC
 AAATTCCGCGGTCTAATGCAGTCACCAATTTCGACATCAGTCGATATGATGTCAAATCA
 ATTGCTAGCTGTAATCTCCCTGTGGGTGGACTAATGCCTAAACCTTCTCCAGCAACCGCA
 CGCGCTGACAAAACCGTTGATCTTCTCCATCCGACTCTCCATCTCTAACCACACCGTCC
 CTCACGTTCAATGTGGCAACACCGGTCAATGACCATGGAGGAACCTTTTACCACACTGGT
 ATACCAATCAAACAGACCCGCTGATCATTATTGGTCCAACATCTTTGGATTCCAGGCA
 AACCCGAAAGCAGAAATGCGACCATTAGCAAACCTTGGGTGGATCTTCATAACCTTCT
 CCTGGTTATGCTATAATGCCGTAATGCAGGAAGGTGAAAACAACCTTGGTGGTAGTTTT
 GTTGGGTCTGATGGGTATAACAATCATTCCGCTGCATCGAACCCGGTCTCAGCAATTCCG
 CTGTCCTCGACAACATAATGAGTAACGGTAACGAAGGGTATGGTGGAAACATAAACTGG
 ATTAATAACAACATTTCAAGTTCTTACCAAACCTGCAAAATCAAATCTCTCTGTTTGCAC

ACACCGGTTTTTGGGTTGGAATGAGTATTCACATCTTAGTGAGAACTAAAATAAATATGT
 AGGAAAAAATAAGGCTCTGTTTGAAGAAATCAGATATTTTCTTCTTAGATTATTTAAGT
 AGTTTAAAAAATATTTTTTAAAGTGTTCACCTTTACGTTTGTCTGCTGACCACGAATT
 TTGCTGGATCTGACAGTACTAACTCTTTGTTTAATGACCTTATGGGTTCCTTTTTACTT
 TCCAGAACTTTTATTTACTTTTTTCTTCATTTTTTCTTCATTTTTTTTGTGTGGGACAAT
 ATGAATGATTGAAGATGGAACTGCTTGCATGTGAATAAACGAAATCAAACNATCTTCG
 GTAACCTAAAAA

>G1009 Amino Acid Sequence (domain in aa coordinates: 201-277, 303-371)

MKNNNKSSSSSSSYDSSLSPSSSSSHQNWLSFSLNNNNNFSSSNPNLTSSTSDHHHP
 HPSHLSLFQAFSTSPVERQDGSFVSPSDATAVLSVYPGGPKLENFLGGGASTTTTRPMQ
 QVQSLGGVVFSSDLQPLHPSPAABEYDSELKSIAASFLGNYSGGHSSEVSSVHKQQPNP
 LAVSEASPTPKKNVESFGQRTSIYRGVTRHRWTGRYEHLWDNSCRREQSRKGRQVYL
 GYDKEDKAARAYDLAALKYWGPTTTTNFPIISNYESELEEMKHMTRQEFVASLRKSSGFS
 RGASMYRGVTRHHQHGRWQARI GRVAGNKDLYLGTSTQEEAAEAYDIAAIKFRGLNAV
 NFDISRYDVKSIASCNLPVGGMLPKPSPATAAADKTVDLSPSDSPSLTTPSLTFNVATPV
 NDHGGTFYHTGIPIKPDADHYWSNIFGFQANPKAEMRPLANFGSDLHNPSPGYAIMPMV
 QEGENNFGGSFVSGDGYNNHSAASNPVSAIPLSSTTTMSNGNEGYGGINWINNNISSY
 QTAKSNLSVLHTPVFGL*

>G170 (1..1107)

ATGGGGATGAAGAAGGTGAAGCTATCTTTGATAGCTAATGAAAGATCAAGGAAAACATCC
 TTCATAAAGAGGAAAGACGGGATTTTTAAGAACTCCACGAGTTGTCAACTCTGTGTGGT
 GTCCAAGCTTGTGCTCTCATCTACAGTCCATTCATACCGGTTCCAGAGTCATGGCCGTCA
 AGGGAAGGTGCTAAAAAGGTGGCTTCAAGGTTTCTGGAGATGCCGCCGACAGCCGAACC
 AAGAAGATGATGGATCAAGAGACTTACCTTATGGAGAGGATTACCAAAGCAAAAGAGCAA
 CTAAGAACCTGGCTGCTGAGAACCGAGAGTTACAGGTTAGACGATTTATGTTTGATTGT
 GTTGAAGGCAAAATGTCCAGTATCATTATGATGCAAAAGACCTTCAAGATTTGCAATCT
 TGTATAAATCTATATCTCGATCAGCTTAACGGAAGGATCGAGTCCATTAAAGAAAATGGT
 GAGTCGTTGTTGTCTTCCGTCTCTCCTTTTCTTACTAGAAATTGGTGTGACGAAATGGT
 GATGAGTCATTTTCCGACTCTCCTATTATGCTACAACCTGGGGTTGTAGATACTCTTAAT
 GCTACCAATCCTCATGTTCTTACGGGCGATATGACTCCTTTTCTTGATGCGGACGCAACT
 GCGGTAACGTCTCCAGTAGATTTTTTGATCATATTCCATATGAAAATATGAATATGAGT
 CAAAATCTGCATGAACCGTTTCAACACCTTGTTCCTACTAACGTTTGTGATTTTTTCAA
 AATCAGAATATGAATCAGGTTCAATACCAGGCTCCTAATAATCTGTTTAATCAGATTCAA
 CGAGAATTCTACAACATAAATTTGAATCTGAATTTGAATCTGAATTCGAATCAGTATCTG
 AATCAACAACAATCATTCATGAATCCGATGGTGAACAACATATGAATCATGTTGGAGGG
 CGTGAAAGCATTCTTTTCGTGGACGGAACTGCTACAACCTACCATCAACTACCATCCAAT
 CAACTACCAGCCGTTGATCATGCTTCCACCAGTTACATGCCTTCCACCACCGGTGTCTAT
 GATCCTTACATCAACAATAATCTCTAA

>G170 Amino Acid Sequence (domain in aa coordinates: 2-57)

MGMKKVKLSLIANERSRKTSFIKRDGIFKKLHELSTLCGVQACALIYSPFIPVSWPS
 REGAKKVASRFLEMPPTARTKKMDQETYLMERITKAKEQLKNLAAENRELQVRRFMFDC
 VEGKMSQYHYDAKDLQDLQSCINLYLDQLNGRIESIKENGESLLSSVSPFPTRIGVDEIG
 DESFSDSPIHATTGVVDLTNATNPVLTGDMTPFLDADATAVTASSRFFDHIPYENMMS
 QNLHEPFOHLVPTNVCDFFQONQNMNQVQYQAPNNLFNQIQREFYNINLNLNLNSNQYL
 NQQQSFMNPMVEQHMHVGGRESIPFVDGNCYNHQLPSNQLPAVDHASTSYMPSTTGVY
 DPYINNNL*

>G1768 (185..1426)

CTTCCTTTTGTCTTCAAGTTCGAGCTTTGGTTGGATCTCTCACTTGCAAAACCAAATCCCT
 TATCGACTTCCACCGAAAGATCACTTCTTAACCTACACAAGGTGTTTGTATGAAGATCA
 GATAAATAAAAGGTCAATTTGAGGATAATGGTTGATGTTCAAAGATTCTTACTTGCTTATT
 TGTGATGGACAATGTAAGAGGTTCAATAATGTTGCAGCCACTGCCAGAGATAGCTGAGAG
 TATCGATGATGCTATCTGCCATGAACCTCTCATGTGGCCTGATGATGCTAAAGATTTGTT
 ATTGATAGTGGAGGCAATATCAAGGGGAGACTTGAAGTTGGTACTTGTTGCTTGTGAAA
 AGCTGTTTCTGAGAATAATCTTCTAATGGCAGGATGGTGTATGGGTGAGTTGCCGCGTAT
 GGTTCGATTTCTGGTGAGCCAATCCAGAGATTGGGAGCTTATATGTTAGAAGGGCTTGT
 TGCTAGGCTTGCTGCTTCTGTTAGTTCGATATATAAGTCTCTCCAGTCCAGAGAACCAGA
 GAGTTATGAATTTTATCTTATGTGTATGTTCTGCATGAGGTTTGTCCATATTTCAAGTT

TGGATACATGTCAGCGAATGGTGCGATTGCAGAAGCAATGAAGGATGAAGAGAGGATTCA
 CATTATTGACTTCCAAATTGGACAAGGGAGCCAGTGGATAGCACTTATCCAGGCTTTTGC
 AGCTAGGCCTGGTGGGGCTCCAAATATTCGAATTACCGGAGTTGGTGATGGATCTGTCTT
 GGTACAGTCAAGAAGAGACTAGAGAACTTGCAAAGAAGTTTGATGTTCCATTACAGGTT
 CAATGCGGTTTCAAGGCCAAGTTGTGAAGTTGAAGTGGAAAATCTTGATGTCCGAGATGG
 CGAAGCCCTTGGAGTGAACCTTTGCTTACATGCTGCATCATTGGCCAGATGAGAGTGTAAG
 CATGGAACACAGGGACCGGTTGCTGAGGATGGTGAAGAGTCTATCACCTAAAGTAGT
 CACTCTTGGAACAAGAATGCAACACGAACACTTCCCCTTCTCCTAGGTTTCTTGA
 GACATTAAGTTATTACACGGCAATGTTTCAATCTATCGATGTTATGCTTCCGAGAAATCA
 CAAGGAAAGGATCAATATCGAGCAGCACTGCATGGCAAGGGATGTCGTCAACATCATAGC
 TTGTGAAGGAGCCGAGAGGATCGAAAGACACGAGCTTCTCGGGAAATGGAAGTCAAGGTT
 TTCCATGGCGGGTTTGTAGCCATACCCCTTGAGCTCAATCATTTCAGCCACCATTAGAGC
 CCTCTTGAGAGATTACAGCAACGGGTATGCGATTGAAGAAAGAGATGGTGCTCTGTACCT
 TGGTTGGATGGACCGAATCTTGGTCTCATCTTGTGCATGGAAGTGAAGAATAAACGTCT
 CCAAGAATGTAATGCAAAAGACAGAAGTGAAGTAATAGATAGTTTTGTCTCATAACCAT
 TAATAAGGTTGAATCAAATCATATACATCCCCATGCTACAACCTATTACACAGGCTCCATC
 AACAAAGAAGGGCTCTTGTGTGTACCTTCTCTTCTGTAACCTTTATTTGAACCAAT
 GGAAGTGGTTACAT

>G1768 Amino Acid Sequence (domain in AA coordinates: 54-413)

MDNVRGSI MLQPLPEIAESIDDAICHEL SMWPDADKDLLLIVEAISRGDLKLVLVACAKA
 VSENLLMARWCMGELRGMVSI SGEPIQRLGAYMLEGLVARLAASGSSYKSLQSREPES
 YEFLSYVYVLHEVCYPYFKFGYMSANGAIAEAMKDEERIHIIDFQIGQGSQWIALIQAFAA
 RPGGAPNIRITGVGDGSLVLTVKKRLEKLAKKFDVFPFRFNAVSRPSCVEVENLDVRDGE
 ALGVNFAYMLHHLDPDESVMENHRDRLLRMVKSLSPKVVTLVEQECNTNTSPFLPRFLET
 LSYYTAMFESIDVMLPRNHKERINIEQHCMARDVNNIACEGAERIERHELLGKWKSRFS
 MAGFEPYPLSSIIISATIRALLRDYSNGYAIEERD GALYLGWMDRILVSSCAWK*

>G185 (77..988)

ATGCAAAAATAAACATAGTAACAATACTTTAAACTATTTACACCACCTTTAATCTTATTCT
 CCACTCTTTGAACGTAATGGAGAAGAACCATAGTAGTGGAGAGTGGGAGAAGATGAAGAA
 CGAGATCAACGAGCTAATGATAGAAGGAAGAGACTATGCACACCAGTTTGGATCAGCTTC
 ATCTCAAGAAACACGTGAACATTTAGCCAAAAAGATTCTTCAATCTTACCACAAGTCTCT
 CACCATCATGAACACTCCGGCGAAGTTGACCAAGTTTCTCAGGGTGGAGGAAGCCCCAA
 GAGCGATGATTCCGATCAAGAACCACTTGTCTCATCAAGAGTTCGAAGAAGTCAATGCCAAG
 GTGGAGTTCAAAAGTCAGAATTGCCCTGGAGCTGGTGTGATAGAACGCTGGACGATGG
 ATTCAAGTTGGAGAAAGTACGGCCAGAAGGATATTCTCGAGGCCAAATTTCCAAGAGGATA
 CTATAGATGCACGTATAGAAAGTCTCAAGGATGTGAAGCCACTAAACAAGTCCAAGATC
 TGATGAAAATCAGATGCTCCTTGAGATCAGTTACCGAGGAATACATTCTTGCTCTCAAGC
 TGCAATGTCCGTACAACATGCCGATACAAAACCTCGAACCAGAACCCAGACCAAGAACA
 CGGAAATCTTGACATGGTAAAGGAAAGTGTAGACAACTACAATCACCAGCACATTTGCA
 TCACAACCTTCACTATCCATTGTCTATCTACCCCAATCTAGAGAATAACAATGCCTATAT
 GCTTCAAATGCGAGATCAAAACATCGAATATTTTGGATCTACGAGCTTCTCTAGTGATCT
 AGGAACTAGTATCAACTACAATTTTCCAGCATCTGGCTCGGCTTCTCACTCAGCATCAAA
 CTCTCCGTCCACCGTCCCTTTGGAATCCCCGTTTGAAGCTATGATCCAAATCATCCATA
 TGGAGGATTTGGTGGGTTCTATTCTTAGTTATCTACTTAAGGGAGGGACGGAACCTTTT
 CATGACCTCTTGATTAAAGAGAGAGTTTTTCATAATAGCTAATCAATTTCTATTCAAATA
 TCCGAGTTTTTTTTCTAATCATGTTTATCAATTGTCTTATTACAGAAGGCTTATTTTCAG
 GTCTATGTTGAAATAAATGGATTGTACTCGTAGGTATGATCCTTGTATTCTAAAAAAA
 AAAAA

>G185 Amino Acid Sequence (domain in AA coordinates: 113-172)

MEKNHSSGEWEKMKNEINELMIEGRDYAHQFGSASSQETREHLAKKILQSYHKSLTIMNY
 SGELDQVSQGGGSPKSDSDQEPLVIKSSKSMRWSKVRIAPGAGVDRTLDDGFSWRK
 YGQKDILGAKFPRGYRCTYRKSQGCEATKQVQRSDENQMLLEISYRGIHSCSQANVGT
 TMPIONLEPNQTOEHGNLDMVKESVDNYNHQAHLHNLHYPLSSTPNLENNAYMLQMRD
 QNIEYFGSTSFSDDLGT SINYNFPASGSASHSASNSPSTVPLESPFESYDPNHPYGGFGG
 FYS*

>G1931 (5..592)

ATCAATGGAAGGGGTTGACAACACAAATCCTATGTTAACCTAGAGAAGGCGAAAACAA

CAATCCTTTTCTTCTTAGATGACAAAACATTAATGATGATGGCTCCTTCGTTAATCTT
 TTCGGGCGATGTAGGTCCATCTTCTTCTTGTACTCCAGCAGGTATCATCTATCTGC
 TCAGCTGGAGAACTTTTCGAGGAGGTGGAGGAGAGATGGGAGGATTAGTGAGTAATAATAG
 CAATAATAGTGATCATAATAAGAATTGCAACAAAGGAAAAGGAAGAGAACTTTGGCAAT
 GCAGAGGATAGCTTTTCATACAAGGAGTGATGATGATGTTCTTGATGATGGTTATCGTTG
 GCGAAAGTACGGTCAGAAATCTGTCAAGAACAATGCTCATCCCAGGAGCTATTATAGATG
 TACATACCACACATGCAACGTGAAGAAACAAGTGCAGAACTGGCAAAAGATCCAAACGT
 TGTGCTAACAACTACGAAGGTGTTTCAATCATCCTTGTGAGAAGCTCATGGAGACTCT
 TAGCCCTCTCCTTAGGCAACTTCAGTTCCTCTCAAGAGTTTCTGATCTGTAATTATTGAA
 TGTTAATTAGTGGTGAATACATTAATTATGCTTTAATCTCTCCATTGACCCTCAATC
 >G1931 Amino Acid Sequence (domain in AA coordinates: 114-170)
 MEGVDNTNPLMTLEEGENNNPFSSLDKTLMMAPSLIFSGDVGPPSSSSCTPAGYHLSAQ
 LENFRGGGGEMGGLVSNNSNNSDNHNKNCNKGKGRITLAMQRIAFHTRSDDDVLDDGYRWR
 KYGQKSVKNNAHPRSYRCTYHTCNVKKQVQRLAKDPNVVVTTYEGVHNHPCEKLMETLS
 PLLRQLQFLSRVSDL*

>G2543 (1..2169)

ATGAGTTTCGTCGTCGGCGTCGGCGGAAGTGGTAGTGGAAGCGGCGGAGACGGTGGTGGT
 AGTCATCATCACGACGGCTCTGAACTGATAGGAAGAAGAAACGTTACCATCGTCACACC
 GCTCAACAGATTCAACGCCTTGAATCGAGTTTCAAGGAGTGTCTCATCCAGATGAGAAA
 CAGAGGAACCAGCTTAGCAGAGAATTGGGTTTGGCTCCAAGACAAATCAAGTTCTGGTTT
 CAGAACAGAAGAAGTCAAGCTTAAAGCTCAACATGAGAGAGCAGATAATAGTGCATAAAG
 GCAGAGAATGATAAAATTCTGTGCGAAAACATTGCTATTAGAGAAGCTCTCAAGCATGCT
 ATATGTCCTAACTGTGGAGGTCTCTCTGTAGTGAAGATCCTTACTTTGATGAACAAAAG
 CTTTCGGATTGAAAATGCACACCTTAGAGAAGAGCTTGAAAGAATGTCTACCATGTCATCA
 AAGTACATGGGAAGACCGATATCGCAACTCTCTACGCTACATCCAATGCACATCTCACC
 TTGGATTGTCAATGACTAGTTTAACTGGTTGTGGACCTTTTGGTCATGGTCTCTCACTC
 GATTTTGATCTTCTTCCAGGAAGTTCTATGGCTGTGGTCTAATAATAATCTGCAATCT
 CAGCCTAACTTGGCTATATCAGACATGGATAAGCCTATTATGACCGGCATTGCTTTGACT
 GCAATGGAAGAATTGCTCAGGCTTCTTCAAGACAAATGAACCTCTATGGACAAGAACAGAT
 GGCTGCAGAGACATTCTCAATCTTGGTAGCTATGAGAATGTTTTCCCAAGATCAAGTAAC
 CGAGGGAAGAACCAGAACTTTCGAGTCGAAGCATCAAGGTCTTCTGGTATTGTCTTCATG
 AATGCTATGGCACTTGTGCGACATGTTTATGGATTGTGTCAAGTGGACAGAAGCTTTTCCC
 TCTATCATTGCAGCTTCTTAAACACTTGCAGTGATTCTTTCAGGAATGGGAGGTACCCAT
 GAGGGTGCATTGCATTGTGTGATGAAGAAATGGAAGTGCTTTTCGCCTTTAGTAGCAACA
 CGCGAATTCTGCGAGCTACGCTATTGTCAACAGACTGAACAAGGAAGCTGGATAGTTGTA
 AACGTCTCATATGATCTTCTCAGTTTGTCTTCTCACTCTCAGTCTTATAGATTTCCATCT
 GGATGCTTGATTTCAGGATATGCCAATGGATATCCAAGGTTACTTGGGTGAACATATT
 GAACTGAAGAAAAGAACTGGTTTCATGAGCTATACAGAGAGATTATTCAGAGGGGATT
 GCTTTTGGGGCTGATCGTTGGGTTACCACTCTCCAGAGAATGTGTGAAAGATTGCTTCT
 CTATCGGTACCAAGCTTCTCATCTCGTGATCTCGGTGGAGTGATTCTATCACCGGAAGGG
 AAGAGAAGCATGATGAGACTTGTCTCAGAGGATGATCAGCAACTACTGTTAAGTGTGAGC
 AGATCCAACAACACACGCTCAACCGTTGTTTCGGAAGTGAACGAAGTTGGAATCCGTGTG
 ACTGCACATAAGAGCCCTGAACCAACCGGCACAGTCTTATGTGCAGCCACCCTTTCTGG
 CTTCCCAATTCTCCTCAAAATGTCTTCAATTTCTCAAGACGAAAGAACCCGTCCTCAG
 TGGGATGTTCTTTCAAACGGAACGAGTGAAGAAGTTGCTCACATCTCAAACGGATCA
 CATCTTGAAACTGCATATCGGTTCTACGTGGATCCAATGCAACACATAGCAACAACATG
 CTTATTCTGCAAGAAAGCTCAACAGACTCATCAGGAGCATTTGTGGTCTACAGTCCAGTG
 GATTTAGCAGCATGGAACATCGCAATGAGCGGTGAAGATCCTTCTTATATTCTCTCTTG
 TCCTCAGGTTTTCACAACTCACCAGATGGAATGGCTCAAACCTGAAACAAGGAGGAGCC
 TCGACGAGCTCAGGACGGGCATCAGCTAGCGGTTGTTGATAACGGTTGGGTTTCAGATA
 ATGGTAAGCAATTTACCGACGGCAAACTGAATATGGAGTGGTGGAAACGGTTAATAAC
 CTGATAGGAACAACGTGATCAAAATTAACCGCCTTGAGCGGTCCTACAGCTTCAACT
 ACAGCTTGA

>G2543 Amino Acid Sequence (domain in AA coordinates: 31-91)
 MSFVVGVGSGSGSGGSGSHHDGSETDRKKRYHRHTAQQIQRLSSFKPCPHPEK
 QRNQLSRELGLAPRQIKFWFQNRRTQLKAQHERADNSALKAEENDKIRCENIAIREALKHA
 ICPNCGGPPVSEDPYFDEQKLRIENAHLEELERMSTIASKYMGRPIQLSTLHPMHISP

LDLSMTSLTGCGPFGHGPSLDFDLLPGSSMAVGPNNNLQSQPNLAISMDKPIMTGIALT
AMEELLRLQLTNEPLWTRTDGCRDILNLGSENVFPRSSNRGKNQNFVEASRSSGIVFM
NAMALVDMFMDVCVKWTELFPSIIAASKTLAVISSGMGGTHEGALHLLYEEMEVLSPLVAT
REFCELRYCQTEQGSWIVNVSYDLPOFVSHSQSYRFPSCGLIQDMPNGYSKVTWVEHI
ETEEKELVHELYREIIHRGIAFGADRWVTTLQRMCFASLSVPASSSRDLGGVILSPEG
KRSMMRLAQRMISNYCLSVSRNNTRSTVSELNEVGIRVTAHKSPEPNGTVLCAATTFW
LPNSPQNVFNFLKDERTRPQWDVLSNGNAVQEVAHISNGSHPGNCISVLRGSNATHSNM
LILQESSTDSSGAFVVYSPVDLAALNIAMSGEDPSYIPLSSGFTISP DGNGSNSEQGGA
STSSGRASASGLITVGFQIMVSNLPTAKLNMESVETVNNLIGTTVHQIKTALSGPTAST
TA*

>G264 (30..1430)

CTTGTACCAGTTTCTGATTAGATTCAACAATGAACGGCGCATTAGGTAACCTCCTCCGCCT
CCGTTAGCGGCGGAGAAGGAGCCGAGGACCGCCTTTCTTGGTGAAAACCTACGAGA
TGGTCGACGATTATCAACGACGACATCGTATCGTGGAGCGCTAACACAACAGCTTCA
TCGTTTGGAAATCATGCCGAATTTTCACGCCTCCTTCTTCCAACCTACTTCAAACACAATA
ACTTCTCTTCTTCAATTCGTGCTCAGCTCAATACCTATGGGTTTAGGAAGATTGATCCAGAGA
GGTGGGAGTTTTTGAATGATGATTTTATTAAGGATCAGAAGCATCTTCTCAAGAATATAC
ATAGAAGGAAACCTATACACAGCCACAGTCATCCACCTGCTTCGTCGACTGATCAAGAAA
GAGCAGTGTGTGAAGAGCAAATGGACAAGCTTTACGTCGAGAAAGCTGCAATTGAAGCTA
AGCTTTTAAAGTTCAAACAACAGAAGGTTGTAGCAAAGCATCAGTTTGAAGAAATGACTG
AGCATGTTGATGATATGGAGAATAGGCAGAAGAAGCTGCTGAATTTTTTGGAACTGCGA
TTCGGAATCCTACTTTTGTGAAGAAATTTGGTAAGAAAGTCGAGCAGTTGGATAITTCAG
CTTACAACAAAAAGCGAAGGCTCCCTGAAGTTGAGCAATCAAAGCCACCTTCAGAAGATT
CTCATCTGGATAATAGTAGTGGTAGCTCGAGACGCGAGTCTGGAAACATTTTTTCATCAAA
ATTTCTCTAATAAATTGCGACTAGAGCTTTCTCCAGCTGATTGAGATATGAACATGGTTT
CACACAGTATACAAAGTTCCAATGAAGAAGGTGCGAGTCCCAAAGGGATACTGTCAGGAG
GTGATCCAAATACTACACTAACAAAAAGAGAAGGCCTACCATTTGCACCTGAAGCTCTAG
AGCTTTCGCGGATACCGGGACATGCCCGAGGAGATTACTGTTAAATGATAATACAAGGGTGG
AGACCTTTCAGCAGAGGCTAACTTCTTCAGAGGAGACTGATGGTAGCTTTTTCATGTCATT
TAAATCTAACCTTGGCTTCTGCTCCGTTACCGGACAAAACAGCTTCACAGATAGCTAAGA
CGACTCTTAAAAGTCAGGAGTTAACTTTAACTCAATAGAAACAAGTGCAAGTGAGAAAA
ATCGGGGTAGACAAGAGATTGCGAGTTGGAGGTAGCCAAGCAAATGCAGCTCCTCCAGCAA
GAGTGAATGATGTATTCTGGGAACAGTTCCCTAACAGAAAGGCCAGGGTCTTCAGATAATG
AGGAGGCAAGTTCGACTTATAGAGGTAACCCATACGAAGAGCAAGAGGAGAAAAGAAACG
GGAGTATGATGTTACGTAATACAAAGAATATCGAGCAGCTGACCTTATAAACTATTTGGA
CGGTTACATCAACGAGAGTACGAAGTGGGTTTGGTAAGAAAGTATGGGTGAGTAAGTAA
TGAAACATTGGACTGAAAAAGCGTAAGTAGCTTTGTTGTAAACACTTGGCTCTCTGTCTA
CACAAGTAATTTGACTGTAAATGTAAGTGTACAGGATTTAAATTGAATAAGCA

>G264 Amino Acid Sequence (domain in AA coordinates: 24-114)

MNGALGNSSASVSGGEGAGGPAPFLVKTYEMVDDSSDQIVSWANNNSFIVWNHAEFSR
LLLPTYFKHNNFSSFIRQLNTYGFRKIDPERWEFLNDDFIKDQKHLLKNIHRRKPIHSHS
HPPASSTDQERAVLQEQMDKLSREKAAIEAKLLKFKQKQVVAKHQFEEMTEHVDDMENRQ
KKLLNPLETAIRNPTFVKNFQKKEQLDISAYNKKRRLPEVEQSKPPSEDSHLDNSSGSS
RRESGNIHFQNFNKLRLLELSPADSDNMVSHSIQSSNEEGASPKGILSGGDPNTTLTKR
EGLPFAPAELELADTGTCPRRLLLNNDNTRVETLQQRLLTSSEETDGSF SCHLNLTLASAPL
PDKTASQIAKTTLLKSQELNFNSIETSASEKNRGRQEIIVGGSQANAAPPARVNDVFWEQF
LTERPGSSDNEEASSTYRGNPYEEQEEKRNKSMMLRNTKNIEQLTL*

>G32 (101..736)

AACACACATTCCTCTCTTCTTCACTAGAAAAAGATAGATATATCGGACATTTATTG
ATCTGTGTATGCAAGGTATAGTATCATTTATTAGAAAGATGAACACAACATCATCAAA
GAGCAAGAAGAAGCAAGACGATCAGGTTGGTACAAGGTTTCTTGGGGTGAGAAGAAGGCC
TTGGGGAAGATACGACGCTGAGATTAGAGACCCAACTACGAAGGAGCGTCACTGGCTTGG
CACTTTCGATACGGCGGAAGAAGCTGCCTTGGCCTACGATAGAGCTGCTCGGTCCATGCG
TGGCACACGTGCCAGAACCAACTTTGTTTACTCAGACATGCCTCCTTCCTCATCCGTCAC
CTCCATTGTTTCTCTGACGATCCTCCTCCTCCTCCACCTCCTCCTGCTCCTCCTAGCAA
TGATCCTGTGCTGATTACATGATGATGTTTAAACCAATACTCATCCACTGACTCGCCAATGCT
TCAGCCTCATTGTGATCAAGTGGACAGTTACATGTTTGGTGGCTCTCAATCTTCGAATTC

TTATTGCTATTCTAATGACAGTAGTAATGAGCTGCCTCCTCTCCCGAGCGACTTGTCGAA
 TTCGTGTTATAGCCAACCACAGTGGACCTGGACCGGTGACGACTACTCGTCTGAGTACGT
 ACATAGTCCAATGTTTCAAGCAATGCCTCCGGTTTCTGACTCTTTCCCTCAAGGTTTCAA
 CTACTTTGGCTCCTAATCTTTCTCATCGTCCATATTTAATACCTTCCTCATTTGTACCT
 TTTCCTTCTTCTCTTTTGGGTTTATCTATGTTTCGCGCTCCTTGATCTCTGCCTATG
 TGATCAAAGTGACTGTTTGTCAATAGTTTTTCAATAACAAGTTATCATTGTATCTTGAA
 AAAAAAAAAA

>G32 Amino Acid Sequence (domain in aa coordinates: 17-84)
 MNTTSSKSKKKQDDQVGRFLGVRRRPWGRYAAEIRDPTTKERHWLGTFTDABEALAYD
 RAARSMRGTRARTNFVYSMDPPSSSVTSIVSPDDPPPPPPPPAPPSPNDPVDYMMMFNQYS
 STDSPMLQPHCDQVDSYMFQGSQSNSYCYNSDSSNELPPLPSDLSNSCYSPQWTWTGD
 DYSSEYVHSPMFSRMPVSDSPQGFNYFGS*

>G436 (1..2157)
 ATGGATTTTACTCGCGATGACAACTCAAGTGATGAACGGGAAAATGATGTAGACGCCAAC
 ACCAACAACCGTCACGAGAAGAAGGGTTACCATCGCCACACTAATGAACAAATTCATAGG
 CTTGAAACGTATTTCAAGGAATGTCCTCATCCAGACGAATTTAGCGACGTCTGTTGGGT
 GAAGAACTGAATCTGAAACCAAAACAAATCAAATTTTGGTTTCAAAACAAAAGAACTCAA
 GCTAAGAGTCACAATGAAAAAGCAGACAATGCAGCGCTTAGGGCAGAAAATATTAAGATT
 AGACGTGAGAACGAATCAATGGAAGATGCACTGAATAATGTGGTTTGCCTCCATGTGGT
 GGTCTGGTCTTGGGAGAGAAGACCAACTTCGACATCTCCAAAACCTCCGTGCACAAAAC
 GCTTATCTCAAAGATGAGTATGAAAGAGTCTCAAACCTACCTAAAACAGTACGGAGGTCAC
 TCAATGCTAATACGTCGAGGCCACACCTATCTCCATGGTCCATCAAACCATGCATCAACG
 TCCAAGAACCGTCCAGCATTGTACGGAACCTCTTCTAACCGTCTCCCGAGCCTTCAAGC
 ATATTTAGAGGACCATACTCGTGGAAACATGAACACCACCGACCGCTCAGCCGCGA
 AAGCCGCTGGAATGCAGAATTTCCAACCACTATCTCAACTGGAGAAAATGCAATGTTG
 GAAGCAGCGGAAAAGCGGTGTGAGAGGTTTTGAGCCTCATTCAAATGGATGATACAATG
 TGGAAAAAGTCGTCTATTGATGATAGGCTCGTCATTGATCCAGGGCTCTATGAGAAATAT
 TTTACTAAGACTAACACAAATGGTCGTCTGAGTCTTCTAAAGATGTCGTGGTGGTTCAA
 ATGGATGCTGGAACCTTGATCGACATCTTCTTAACTGCGGAGAAATGGGCGAGGCTTTTT
 CCAACAAATGGAACGAAGCTAAAACGATTACGCTCTGGATTCCGTTGACCATCGAGGA
 AAAACTTTCTCAAGAGTGATTATGAGCAACTGCACATACTGTCAACATTGGTGCCACCG
 AGGGAATTTATGATCCTAAGGACTTGCCAACAAATGAAGACAATGTCTGGATGATTGCT
 GATGTGTGCTGTCATCTCCCAACATTGAGTTTGATCTTTCTGTTTCCCATTTGCACCAA
 CGTCCCTCAGGTGTGCTCATTCAAGCCTTGCCCCACGGCTTCTCTAAGGTGACGTGGATA
 GAGCATGTGGTAGTGAATGATAATAGAGTGCAGGCCACATAAGCTTTACAGAGACCTCTTA
 TACGGCGGCTTTGGCTACGGAGCTCGACGTTGGACCGTTACTCTTGAGAGGACGTGTGAG
 AGGCTGATTTTCTCCACCTCCGTCCCTGCCTTGCCCAACAATGACAATCCCGGAGTTGTG
 CAAACAATACGAGGCAGAAATAGCGTAATGCATTTGGGAGAAAGATGTTGAGGAACCTT
 GCATGGATGATGAAAATGTTTAAACAACTCGACTTCTCGCCACAGTCTGAAACTAACAAC
 AGCGGAATTAGGATTGGGGTGGCGATAAACAATGAGGCGGGTCAACCGCCCGTCTCATT
 GTCTGTGCTGGTTTCATCTTTATCCCTCCCTCTCCCTCCTGTCCAAGTGACGATTTCTT
 AAGAATCTGGAGGTTCTGACCAAGTGGGACGTTCTGTGCCATGGGAATCCAGCGACTGAG
 GCTGCTCGTTTCGTACCCGATCAAACCCAAGGAACACTGTGTCTTTCTCGAGCCTTCA
 ATTAGGGATATTAATACTAAGCTAATGATACTCAAGATAGCTTCAAAGATGCATTGGGA
 GGAATGGTGGCCTACGCTCCAATGGATCTAAACACCGCCTGCGCTGCCATTTTCAGGCGAT
 ATCGATCCTACCACCATTTCAATCCTCCCTTCCGGTTTTATGATCTCCCGTGACGGCCGT
 CCTTCCGAGGCGAAGCGAGGGTGGCAGCTATACACTCCTCACCCTGGCTTTCCAGATC
 CTTGTCTCCGGTCCSAGTTACTCTCTGATACCAACCTGGAAGTTCTGTCACACACAGTC
 AATACCTTGATTAGCTCCACCGTTCAAAGGATCAAAGCCATGCTCAAGTGCGAATGA

>G436 Amino Acid Sequence (domain in AA coordinates: 22-85)
 MDFTRDDNSSDERENDVDANTNNRHEKKGYHRHTNEQIHRLETYFKECPHPDEFQRRLLG
 EELNLKPKQIKFWFQNKRTQAKSHNEKADNAALRAENIKIRRENESMEDALNNVCPPCG
 GRGPGREDQLRHLQKLRAQNAYLKDEYERVSNYLKQYGGHSMHNVEATPYLHGPNHAST
 SKNRPALYGTSSNRLPEPSSIFRGPYTRGNMNTAPPQPRKPLEMQNFQPLSQLEKIAML
 EAAEKAVSEVLSLIQMDDTMWKKSSIDRLVIDPGLYEKYFTKNTNNGRPESKDVVVVQ
 MDAGNLIDIFLTAEKWARLFPTIVNEAKTIHVLDSVDHRGKTFSRVIYBQLHLISPLVPP
 REFMLRTCQQIEDNVWMIADVSHLPNIEFDLSFPICTKRPSGVLIQALPHGFSKVTWI

EHVVVNDNRVRPHKLYRDLLYGGFGYGARRWTVTLERTCERLIFSTSVPALEPNNDNPGVV
 QTIRGRNSVMHLGERMLRNFAMMMKMNKLDSPQSETNNSGIRIGVRINNEAGQPPGLI
 VCAGSSLSLPLPPVQVYDFLKNLEVRHQWDLVCHGNPATEAARFVTGSNPRNTVSFLEPS
 IRDINTKLMILQDSFKDALGGMVAYAPMDLNTACAAISGDIDPTTIPILPSGFMISRDGR
 PSEGEAEGGSYTLTLTVAQILVSGPSYSPDTNLEVSATTVNTLISSTVQRIKAMLKCE*
 >G556 (50..1144)

CTTTTTTGAAGCCCTTTTGACACAAAAGACCAGAACAAAGTTGAAGAAATATGAATACAAC
 CTCGACACATTTTGTTCACCGAGAAGGTTTGAAGTTTACGAGCCTCTCAACCAAATCGG
 TATGTGGGAAGAAAGTTTCAAGAACAATGGAGACATGTATACGCCTGGCTCTATCATAAT
 CCCGACTAACGAAAAACCAGACAGCTTGTCTAGAGGATACTTCTCATGGGACAGAAGGAAC
 TCCTCACAAAGTTTGACCAAGAGGCTTCCACATCTAGACATCCTGATAAGATACAGAGAAG
 GCTAGCACAGAATCGAGAGGCAGCTAGGAAAAGTCGTTTGCAGCAAGAAAGCTTATGTTCA
 GCAGCTAGAGACTAGCCGGTTAAAGCTAATTCATTTAGAGCAAGAACTCGATCGTGCTAG
 ACAACAGGGTTTCTATGTGGGGAACGGAGTAGATACCAATGCTCTTAGTTTCTCAGATAA
 CATGAGCTCAGGGATTGTTGCATTTGAGATGGAATATGGACATTGGGTGGAAGAACAGAA
 CAGGCAAAATATGTGAACATAAGAACGGTTTTACATGGACAAGTTAGTGATATAGAGCTTCG
 TTCTCTAGTCGAGAATGCCATGAAACATTACTTTCAACTCTTCCGAATGAAGTCAGCCGC
 TGCAAAAATCGATGTTTTCTATGTCATGTCCGGAATGTGGAAAACCTCAGCAGAGCGGTT
 TTTCTTGTGGATAGGCGGATTTAGACCCTCAGAGCTTCTCAAGGTTCTGTTACCGCATT
 TGATCCTTTGACGGATCAACAACCTTTGGATGTATGTAATCTGAGGCAATCATGTCAACA
 ATCAGAAGATGCGTTATCCCAAGGTATGGAGAACTGCAACATACATTAGCAGAGAGTGT
 AGCAGCCGGGAAACTTGGTGAAGGAAGTTATATTCCTCAAATGACTTGTGCTATGGAGAG
 ATTGGAGGCTTTGGTCAGCTTTGTAAATCAAGCTGATCATCTGAGACATGAGACATTGCA
 ACAGATGCATCGGATCTTAACCACGCGACAAGCGGCTAGAGGTTTGTAGCATTAGGGGA
 GTATTTCCAAAGGCTTCGAGCTTTGAGTTGAGTTGGGCGGCTAGGCAACGTGAACCAAC
 GTAATTAAGGTGTTTAGATGTCAAGAAAGGTTTGAGACCTTAACAATCAAGAAATGGAGTT
 TGCTGGTGAGTGGATTTTTGGGTCAAGAACAGAGCAATAACACAAGCTGCTGTGTGATG
 ATGAATCTTGTCTTGCGGCTAAAGGAAATGTTTGAGGAAAGTTGTACATATGATCAGCAA
 CGTAAAGTTTATAGCTTTTGTAGAAACCAACTTTTCGATGGTTGTTCTTTTTTTTTTGTAT
 GTAATATTATAGATAAGCTTGTGGTATATATGATTTTAATGTGACATTACGAACCTTGATT
 TATAACCATGGTAAAAT

>G556 Amino Acid Sequence (domain in AA coordinates: 83-143)
 MNTTSTHFVPPRRFEVYEPLNQIGMWEESEFKNNGDMYTPGSIIIPTNEKPDSLSEDTSHG
 TEGTPHKFDQEAESTSRHPDKIQRRLAQNREARKSRLRKKAYVQOLETSRLKLIHLEQEL
 DRARQQGFYVNGVDNLSFSDNMSSGIVAFEMEYGHWWVEEQNRQICELRTVLHGQVSD
 IELRSLVENAMKHYFQLFRMKSAKIDVFYVMSGMWKTSARFFLWIGGFRPSELLKVL
 LPHFDPLTDQQLLDVCLNRQSCQSEDALSQGMELQHTLAESVAAGKLGEGSYIPQMT
 C
 AMERLEALVSFVNQADHLRHETLQQMHRILTTTQAARGLLALGEYFQRLRALSSSSWAARQ
 REPT*

>G1420 (39..1238)

AAAGTATCATCTCATAGATTCCATCTTTTCTCTATTACATGGAGAAGAAAAAGAGAGG
 ATCATCATCATCAACAACAACAACAACAAGGAGATCAAGAACACAGAGACAAAGA
 TCGAGCAAGAACAAGAACAAGAAAAACAAGAAATCTCTCAAGCATCATCATCATCAA
 ACATGGCGAATCTAGTTACGTCATCAGATCATCATCCGTTGGAGCTAGCTGGAATCTCT
 CAAGCATCTTCGATACTTCATCTTTACCTTTTCCCTTATCTTATTTTGAAGATCACTCTT
 CTAATAATCCTAATTCTTTCCCTAGACTTGCTCCGACAAGATCATCAGTTTGCTTCTTCT
 CTAATTCCTCTTCTTTTTCATTTCGATGCCTTTCTCTCCCCAATAACAACAACAACCT
 CTTTTTTTACGGATTGCCCCTTACCTCAAGCTGAGTCATCAGAAGTCGTGAACACAACAC
 CGACTTCTCCAAACTCAACCTCAGTCTCATCTTCTCCCAACGAAGCTGCAATGATAACA
 ACAGTGGTAAAGAAGTTACTGTTAAAGATCAAGAAGAAGGAGATCAACAACAAGAGCAAA
 AGGGTACTAAGCCACAGTTGAAGGCAAGAAGAAGAAATCAAAAGAAAGCTAGAGAAGCTA
 GGTGTTGCGTTTCTGACGAAGAGCGATATTGATAATCTTGACGACGGTTATAGGTGGAGAA
 AATACGGCCAAAAAGCTGTCAAAAACAGTCCTTATCCAGAAGCTATTACCGTTGCACCA
 CAGTGGGTTGCGGAGTGAAGAAGAGAGTGGAGAGATCTCCGATGATCCTTCGATCGTCA
 TGACAACCTACGAAGGTCAGCATACCCATCCTTTCCCCATGACGCCACGTGGACACATCG
 GAATGCTCAGGTCACCAATCCTAGACCACGGTGCAACCACCGGTCATCATCATCATCT
 CCATCCCTCAGCCACGTTACTTGCTGACTCAACATCACCAGCCCTACAACATGTACAACA

ACAACTCTCTAAGTATGATCAATAGAAGATCATCCGATGGCACTTTCGTAAATCCAGGTC
CATCATCATCATTCCTCCCGGCTTTGGTTATGATATGTCTCAAGCTTCTACTTCAACTTCTT
CTTCCATTAGAGATCATGGATTGCTTCAAGATATCTTCTTCGCAGATCAGATCCGATA
CTATTAACACTCAAACCAATGAAGAGAATAAGAAATGAAGAAGTTTTTTTCCCGGGCA
ATTGTTTTTTTCTTTAGGCCGGATCCGGTAGGTAGGTTTCATGAGC
>G1420 Amino Acid Sequence (domain in AA coordinates: 221-280)
MEKKKEEDHHHQQQQQQKEIKNTETKIEQEQEKEISQASSSSNMANLVTSSDHP
LELAGNLSIFDTSLLPFPYSYFEDHSSNNPNSFLDLLRQDHQFASSNSSSSFSFDAFPL
PNNNNNTSFFTDLPLPQAESSEVVNTTPTSPNSTSVSSSSNEAANDNNSGKEVTVKDQEE
GDQQQEOKGTPQLKAKKKKQKAREARFAFLTKSDIDNLDDGYRWRKYGQKAVKNSPYP
RSYYRCTTVGCGVKRVERSSDDPSIVMTTYEGQHTHPFPMTPRGHIGMLTSPILDHGAT
TASSSSFSIPQPRYLLTQHHQPYNNMNNNSLSMINRRSSDGTFFVNPGPSFPGFYDMS
QASTSTSSSIRDHGLLDILPSQIRSDTINTQTNEENKK*
>G1412 (115..1008)
CCCACGCGTCCGCCACGCGTCCGAAACAAAAACATATAATTTGGGTTTTTAGAGTTCGA
AACTTGAAATCTTTTTTTTTTTGGTTGCTGAGGAATCGAAGTAGAAGAGTATAAATGGGT
GTTAGAGAGAAAGATCCGTTAGCCAGTTGAGTTGCCACCAGGTTTTAGATTTTATCCG
ACAGATGAAGAGCTTCTTGTTTCAATATCTATGTCGGAAGTTGCAGGCTATCATTTCTCT
CTCCAGGTCATCGGAGACATCGATCTCTACAAGTTCGATCCTTGGGATTGCGCAAGTAAG
GCTTTGTTTGGAGAGAAGGAATGGTATTTCTTTAGCCCCAAGAGATCGGAAATATCCGAAC
GGGTCAAGACCCAATAGAGTAGCCGGGTCGGGTATTGGAAAGCAACGGGTACTGACAAA
ATTATCACGGCGGATGGTCGTGTCGGGATTAAAAAGCTCTGGTCTTTTACGCCGGA
AAAGCTCCCAAAGGCATAAAACCAACTGGATTATGCACGAGTATCGCTTAATAGAACAT
TCTCGTAGCCATGGAAGCTCCAAGTTGGATGATTGGGTGTTGTGTGCAATTTACAAGAAA
ACATCTGGATCTCAGAGACAAGCTGTTACTCCTGTTCAAGCTTGTCTGTAAGAGCATAGC
ACGAATGGGTCGTATCTCTTTCATCACAGCTTGACGACGTTCTTGATTGCTTCCCG
GAGATAAAAGACCAGTCTTTTAATCTTCTCGGATGAATTCGCTCAGGACGATTCTTAAC
GGGAACTTTGATTGGGCTAGCTTGGCAGGCTTTAATCCAATTCAGAGCTAGCTCCGACC
AATGGATTACCGAGTTACGGTGGTTACGATGCGTTTCGAGCGGCGGAAGGTGAGGCGGAG
AGTGGGCATGTGAATCGGCAGCAGAACTCGAGCGGTTGACTCAGAGTTTCGGGTACAGC
TCGAGTGGGTTTGGTGTTCGGGTCAAACATTGAGTTTAGGCAATGAGAGAGATGTGAA
GTTACTGATGGGTGAAAAAAGTAAAAAATACTGGAGATAGTAGAGTGGCAATTGATG
TAAATAATAGGGATTTATATGGGGCTTTTACCGATTGCGGTGAGGCTTAGGATTCCCCAAA
GGAAAAAGGCTCGACTGGGGACTAGTTTGATCCAACCTTGACGGCCCCCAATGTGTAATG
TTTCTCAACGGAGAGAAAAATAATGGTTACCAATATTTTCCAAAAAATAAATAAATAA
>G1412 Amino Acid Sequence (domain in AA coordinates: 17-159)
MGVREKDLAQLSLPPGFRFYPTDEELLVQYLCKRVAGYHFSLVQVIGDIDLYKFDPWDL
SKALFGEKEWYFFSPRDRKYPNGSRPNRVAGSGYWKATGTDKIIADGRRVGIKKALV
AGKAPKGTKNWIMHEYRLIEHSRSHGSSKLDWVLCRIYKKTSGSQRQAVTPVQACREE
HSTNGSSSSSSQLDDVLDSPFIKQSFNLPRMNSLRITLNGNFDWASLAGLNPIPELA
PTNGLPSYGGYDAFRAEAGEAESGHVNRQONSSGLTQSFYSSSGFVSGQTFEFRQ*
>G738 (1..885)
ATGGACCATCATCAGTATCATCATCATGATCAATACCAACATCAGATGATGACTAGTACT
ACAATAATTCCTATAACACCATCGTCACAACACAACCACCACCAACAACAACAATG
GATTCAACAACAGCAACAACATATGATAATGGATGACGAGAAGAAGTTGATGACGACAATG
AGCACTAGGCCGCAAGAACCAAGAACTGTCCAAGATGCAACTCAAGCAACACCAAGTTT
TGTTATTACAACAACATACAGCTTAGCACAGCCTAGGTACTTGTGTAAGTCTTGTGCGGAG
TATTGGACTGAAGTGGCTCTCTCCGTAACGTCCCGTAGGCGGAGGTTCTAGAAAGAAC
AAGAAGCTTCCATTTCCTAATTCCTCTACTTCTTCTCCACCAAGAACCTCCCGGATCTC
AACCCTCCTTTCTGTTTACATCATCAGCTTCATCATCAAACCCTAGCAAGACGCATCAA
ACAATAATGACCTCAGCCTATCCTTCTCCTCCCTATGCAAGACAAGCGAGCTCAAGGG
CATTACGGTCATTTTCAAGTGAAGTTGTGACAGGAGGGCAGAACTGTCTTTTCCAAGCT
CCTATGGGAATGATTCAAGTTTCTGTAAGAGTATGATCATGAGACCCCAAAAAGAATCTT
GGGTTTTTCATTAGACAGGAACGAGGAAGAGATTGGTAATCATGATAACTTCGTTGTTAAT
GAGGAAGGAAGTAAGATGATGTATCCTTATGGAGATCATGAAGACCGTCAACAACATCAC
CATGTGAGACACGATGATGGTAATAAGAAGAGAGAAGGTGGTTCAAGCAATGAGCTATGG
AGCGGAATCATCCTAGGTGGTGTAGTGGTGGACCAACATGGTGA

>G738 Amino Acid Sequence (domain in aa coordinates: 351-393)
 MDHHQYHHHDQYQHMMSTNNNSYNTIVTTQPPPTTTTMDSTTATTMIMDDEKKLMTTM
 STRPQEPNRCPRCNSSNTKFCYNNYSLAQPRYLCKSCRRYWTEGSLRNVPVGGGSRKN
 KKLPPFNSSTSSSTKNLPLNPPFVFTSSASSSNPSKTHQNNNDLSLSFSSPMQDKRAQG
 HYGHFSEQVVTGGQNCFLQAPMGMIQFRQEYDHEHPKKNLGFSLDRNEEEIGNHDFVFN
 BEGSKMYPYGDHEDRQHHHVRHDDGNKKREGGSSNELWSGIILGGDSGGPTW*

>G2426 (1..1038)

ATGGGCAGATCGCCATGTTGTGATAAGGCCGGGTTGAAGAAAGGCCCTTGACTCCAGAA
 GAGGATCAGAACTTTTGGCTTATATTGAAGAACATGGCCATGGAAGCTGGCGTTCTTTG
 CCTGAGAAAGCCGGTCTCCAAAGGTGTGGAAAGAGTTGCAGACTCAGATGGACTAACTAC
 CTAAGACCTGACATCAAGAGAGGCCAAATTCAGTGTACAAGAAGAACAAACCATCATTCAA
 CTCCACGCTCTCCTCGGAAACAGGTGGTCAGCGATTGCAACTCATTTACCAAAGAGGACA
 GACAACGAGATCAAGAACTACTGGAACACACACTTGAAGAAACGTCTGATCAAAATGGGG
 ATAGATCCAGTGACTCACAAAGCACAAAAACGAGACTCTTTCTGCTTCCACAGGACAATCA
 AAGAACGCGAGCCAGCTTAGTCATATGGCTCAATGGGAGAGTGCAAGACTCGACGCTGAA
 GCAAGGCTAGCTAGAGAAATCAAAGCTTCTCCATTTACAGCATTACCAAACAATAACAAC
 CTTAACAATCAGCAGCTCCTCAACAACATTGCTTCACTCAAAAAACATCAACAACTGG
 ACTAAACCAAACCAAGGAAACGGAGACCAACAGCTTGAATCTCCGACATCGACGGTGACA
 TTCTCTGAGAATCTTCTGATGCCTTTAGGAATCCCTACGGATAGCAGCAGAAATAGAAAC
 AATAACAACAATGAGTCTCGGCGATGATTGAATTGGCCGTATCTTCGTCAACCTCCTCC
 GATGTGAGTCTGGTCAAAGAACATGAACACGACTGGATTAGGCAGATCAACTGTGGTAGT
 GGAGGAATAGGAGAAGGATTACGAGTCTATGATCGGTGATTGGTTCGGCCGGGGTTTA
 CCCACCGGGAAAAACGAAGCGACGGCGGGCTGGGGAATGAGAGTGAGTATACTACTAT
 GAGGATAACAAGAATTACTGGAATAGCATTCTCAACTTGGTTGATTCTTCACCGTCCGAT
 TCCGCGACGATGTTCTGA

>G2426 Amino Acid Sequence (conserved domain in AA coordinates:14-114)

MGRSPCCDKAGLKKGPWTPEEDQKLLAYIEEHGHSWRSLEKAGLQRCGKSCRLRWNTY
 LRPDIKRGKFTVQEEQTIQLHALLGNRWSAIATHLPKRTDNEIKNYWNTHLKKRLIKMG
 IDPVTHKHKNETLSSSTGQSKNAATLSHMAQWESARLDAEARLARESLLHLQHYQNNNN
 LNKSAAPQQHCFTQKTSTNWTNPKPNQNGDQQLSPTSTVTFSENLLMPLGIPTDSSNRN
 NNNNESSAMIELAVSSSTSDVSLVKEHEHDWIRQINCGSGGIGEGFTSLIGDSVGRGL
 PTGKNEATAGVGNESYNYEDNKNYWNISILNLVDSSPSDSATMF*

>G1524 (1..825)

ATGGGGAGAACTAAGGAGCAGGCAACATTAACCTCGGTATCCACCCTGTCTTAGGAATCCT
 GCTAAATTCAATGATATAAACAAGCACTCCAGGAAAAAGGATATGGTAAGGCTCTGAAA
 AGAAAACCTTGACCGGTGTGACATGCCCTGTCTGTCTTGAGGTTCTCACAACCTCGGT
 GTCCTCCTTTGTTTCATCTTACCACAAAGGATGCCGTCCGTACATGTGTGCCACGGGAAAC
 CGTTTCTCAAATTGTCTAGAGCAGTACAAAAAGGCATATGCCAAGGATGAGAAAAGTGAC
 AAACCGCCAGAGCTATTGTGCCCGCTTTGTAGGGGTGAGGTGAAAGGCTGGACCGTTGTG
 GAAAAGGAACGTAAGTATCTGAATTCTAAGAAAAGGTCAATGATGAACGACGAGTGTG
 TTTTATGGAAGCTATAGACAGCTCAAGAAGCATGTTAAGGAGAACCATCCGAGAGCCAAG
 CCAAGAGCCATAGACCCTGTGCTGGAGGCGAAATGGAAGAAGCTTGAGGTTGAGAGGGAG
 AGGAGTGATGTAATCAGCACAGTCATGTCTCAACACCTGGGGCTATGGTATTTGGAGAC
 TATGTGATTGAGCCATACAATGGTTATGATCATCAAGATGACAGTGACGATTACAGTGAT
 TCGTCCGATGACGAAATGGAAGGTGGGGTATTCGAGCTTGAGCATTGACCTGGGCCGT
 CTTCAACCGCGTTTCGGCTGCCATCTCAAGCCGGGGAATTCGCGGTATGATCATAAGGAAC
 CGGTGGGCTCGAAGCAGAGGTGCGAGCAGAAGGCGACAAACATAA

>G1524 Amino Acid Sequence (conserved domain in AA coordinates:49-110)

MGRTEQATLTRYPPCPNPAKFNDINKALQEKYKALKRKPWTGVTCPVCLVPHNSV
 VLLCSSYHKGCRPYMCATGNRFSNCLEQYKAYAKDEKSDKPELLCPLCRGQVKGWTVV
 EKERYLNSKKRSCMNDECLFYGSYRQLKKHVKENHPRAKPRAIDPVLEAKWKLEVERE
 RSDVISTVMSSTPGAMVFGDYVIEPYNGYDHQDDSDSDSSDDEMBGGVFELGAFDLGR
 LQPRSAAISSRGIRGMIIRNRWARSRGASRRRQT*

>G1243 (1..3174)

ATGGCGAGAAATTGCAATTCCGATGAGGCTTTCTCGTCAGAGGAGGAAGAAGAGCGGGTT
 AAGGATAATGAAGAAGAAGATGAGGAGGAGCTCGAGGCTGTTGCTCGTCTTCTGGCTCC
 GACGATGACGAAGTAGCCGCCGCCGACGAATCACCAGTCTCCGACGAGAGGCTGCTCCC

GTAGAAGATGATTACGAGGACGAAGAAGATGAGGAAAAAGCTGAAATCAGCAAACGTGAG
AAAGCCAGACTTAAAGAGATGCAGAAGTTGAAGAAGCAGAAGATTCAAGAGATGCTGGAG
TCGCAGAATGCTTCCATTGACGCGGATATGAACAATAAGGGAAAAGGGAGACTGAAGTAT
CTTCTGCAGCAAACCTGAGTTATTTGCCCACTTTGCTAAAAGTGATGGATCTTCTTCTCAG
AAGAAGGCAAAAGGAAGAGGATGCTTCCAAAATAACTGAAGAGGAGGAAGACGAA
GAGTATCTAAAGGAAGAAGAGGATGGCTTAACTGGATCTGGAACACACGTTACTCACA
CAGCCCTCTTGATTCAAGGGAAGATGAGAGATTACCAATTAGCTGGTTTGAAGTGGCTC
ATTCGTCTTTATGAGAATGGCATAAATGGAATTCTTGCTGATGAAATGGGTCTGGGGAAG
ACGCTTCAAACGATTTCTTTGTTGGCATATCTTCATGAATACAGGGGAATCAATGGTCCC
CATATGGTGGTTGCTCCAAAATCAACACTTGGTAATTGGATGAACGAAATTCGCCGTTT
TGTCCTGTCTACGTGCTGTGAAGTTCCTTGGTAATCCTGAGGAGAGGAGACATATTCGA
GAAGACCTGCTAGTTGCTGGGAAATTTGATATTTGTGTCACAAGCTTTGAGATGGCCATC
AAAGAGAAGACAGCACTTCGTCGGTTTAGCTGGCGTTATATTATCATTGATGAAGCGCAT
CGAATCAAGAACGAGAATTCACTCCTTTCTAAAACCATGAGACTTTTGTAGCACCATTAT
CGGCTTCTTATCACGGGGACCCCTTCAGAATAATCTCCATGAAGTGTGGGCTCTTCTA
AATTTTCTTCTGCTGAGATTTTGTAGTTCAGCAGAGACTTTTGATGAATGGTTTCAAATT
TCTGGTGAGAATGACCAGCAAGAAGTTGTGCAACAACCTGCACAAGGTTCTTCGACCATT
CTTCTTCAAGACTAAAGTCAGATGTTGAGAAAGGTTTGCACCGAAGAAGGAGACCATA
CTTAAAGTTGGTATGTCTCAGATGCAAAAGCAATACTACAAGGCTTTACTGCAGAAGGAT
CTTGAAGCGGTTAATGCTGGTGGAGAACGCAACGCTCTGCTAAACATTGCAATGCAACTG
CGTAAATGCTGCAATCACCCCTATCTCTTCCAGGGTGCAGAACCTGGTCCCCCATATACC
ACAGGAGATCACCTTATAACAAATGCTGGTAAGATGGTTCTCTTGGATAAATTGCTTCT
AAGTTGAAAGAACGTGATTCAAGGGTGTCTGATATTTTCTCAGATGACAAGACTTTTGAT
ATTCTTGAGGACTATTTAATGTATCGTGGTTACTTGTATTGCCGTATTGATGAAACACT
GGTGGTGACGAACGAGATGCCTCCATAGAAGCCTACAACAAGCCAGGAAGTGAGAAATTT
GTTTTCTTGTATCTACTAGAGCTGGAGGGCTTGGTATCAATCTTGCTACTGCAGATGTT
GTGATCCTTTACGATAGTGATTGGAACCCACAAGTCGACTTGCAAGCTCAGGATCGTGCC
CATAGGATTGGTCAAAAAAAGAAGTTCAAGTGTTCGATCTGCACTGAGTCTGCTATT
GAGGAGAAAGTGATTGAAAGAGCTTACAAGAAGTTAGCACTTGATGCTCTGGTTATTCAA
CAAGGGAGATTGGCAGAACAGAAAAGTAAGTCTGTCAATAAGGATGAGTTGCTTCAAATG
GTAAGATATGGTGTCTGAGATGGTGTTCAGTTCTAAAGATAGCACAATCACAGACGAGGAT
ATTGATAGAAATCATTGCCAAAGGAGAGAGGCAACAGCTGAACCTTGATGCTAAGATGAAG
AAATTCACAGAAGATGCTATACAGTTTAAATGGATGACAGTGCTGACTTCTATGATTTT
GATGATGACAATAAGGATGAAAACAAGCTCGATTTTAAAAAGATTGTAAGCGACAATTGG
AATGATCCCCCAAGCGGGAGAGAAAGCGCAACTACTCTGAATCTGAGTACTTTAAGCAA
ACATTGCGGCAAGGTGCTCCAGCTAAACCTAAAGAGCCTAGAATTCCGCGCATGCCCCAG
TTGCACGATTTCCAGTTCTTTAACATTGAGAGATTGACCGAGTTGTATGAAAAGGAAGTA
CGTTATCTCATGCAACACATCAGAAAAATCAGTTGAAAGACACAATTGATGTTGAAGAA
CCAGAAGTTGGGATCCCTTAACTACTGAAGAAGTAGAAGAAAAGGAGGGATTATTGGAG
GAGGGTTTCTCAACATGGAGCAGAAGAGATTTTAACTTTCTCAGGGCTTGTGAGAAG
TATGGCCGCAACGACATAAAAAGCATTGCCTCTGAGATGGAAGGGAAAACAGAGGAAGAA
GTTGAAAGATATGCCAAAGTATTTAAAGAGCGGTACAAGGAGCTGAACGACTATGATAGA
ATCATTAAAGAACATTGAGAGGGGAGAGGCAAGGATCTCTAGGAAAGACGAAATCATGAAG
GCCATAGGGAAGAACTGGATCGCTACAGAAACCTTGGCTGGAAGTGAAGATTCAATAT
GGTCAGAACAAAGGCAAGCTGTACAATGAAGAGTGTGACCGTTTCATGATCTGCATGATT
CACAACTTGGTTATGGGAATTGGGATGAGCTAAAGGCAGCATTTAGGACATCGTCTGTG
TTCAGTTTGACTGGTTTGTGAAATCCCGCACGAGTCAGGAACCTTGCAAGAAGATGCGAC
ACTCTGATTGCTGATCGAGAAAGAGAACAGGAGTTTGATGAAAGAGAGAGGCAAGCC
CGCAAAGAGAAGAAGCTCGCGAAGAGTGCAACACCATCAAAGCGACCTTTAGGAAGACAA
GCAAGTGAGAGTCTTTCATCGACGAAGAAGCGGAAGCACCTGTGATGAGATGA
>G1243 Amino Acid Sequence (domain in AA coordinates: 216-609)
MARNSNSDEAFSSEEEEEERVKDNEEEDDEEELEAVARSSGSDDEVAADSPVSDGEAAP
VEDDYEDEEDEEKAEISKREKARKEMQKLKKQKIQEMLESQNASIDADMNNKGKGRLLKY
LLQQTLELFAHFAKSDGSSSQKAKGRGRHASKITEEEDEEYLKEEEDGLTGSGNTRLLT
QPSCIQKMRDYQLAGLNWLIRLYENGINGILADEMGLKTLQTISSLAYLHEYRGINGP
HMOVAPKSTLGNWMEIRRFPCVLRVAVKFLGNPEERRHIREDLLVAGKFDICVTSFEMAI
KEKTALRRFSWRYIIIDEAHRIKENSLLSKTMRLFSTNYRLLITGTPLQNNLHELWALL

ACTCAACTCGGAGGTATTGGTTTGAACCTAGCCGCTACTAATGGCAACAACCAAGCTCAC
CAGATCGGTTCCAGTTTGATGATGAGCGATCTAGGGTTTCTCCATGGACGAAATACTTCA
ACTCCGATGACGGGAAACATTTCATGAAAAACAACAATAATAACAATGAAAAACAACCTA
ATGGCATCCGTTGGATCTTTGAGCCCTTTGCTCTCTTCGATCCAACGACGGGGCTATAC
GCTTTCCAGAACGACGGTAATATCGGGAACAACGTTGGGATATCTGGTTCTTCTACTTCC
ATGGTTGATTCTAGGGTTTATCAGACGCCTCCGGTGAAGATGGAAGAACAACCTAATTG
GCTAATTGTCTAGACCGGTCTCCGGTTTGACGTCTCTGGGAATCAAACAATCAGTAC
TTTTGGCCTGGTTCCGATTTCTCGGGTCTTCTAATGATCTCTTGTGA
>G1909 Amino Acid Sequence (conserved domain in AA coordinates:23-51)
MGSMAERARQANIPPLAGPLKCPKCDSSNTKFCYNNYNLTQPRHFCKGCRRYWTQGGG
LRNVVPVGGGCRNRNKKGNLKSSSSSSKQSSSVNAQSPSSGQLRTHNQFPFSPTLYNL
TQLGGIGLNLAAATNGNNQAHQIGSSLMSDLGFLHGRNTSTPMTGNIHENNNNNNNENNL
MASVGSLSPPALFDPTTGLYAFQNDGNIGNNVGISGSSSTMVDSRVYQTPPVKMEEQPNL
ANLSRPVSGLTSPGNQTNQYFWPGSDFSGPSNDLL*
>G1663 (64..630)
TTCTCTCTGTGAATCCTTGTTTCATCGTCACTGAAATTAGTTTACAAAATCGACGAATTTCG
GAGATGATTTTTTCAAGATGTGTGCAGAAATGAGTCCAACCTTCAACGCTATAGCTTCCGAA
TCGCGTTCCCAAACGCAGTTCCGTGTTTCGAAATCCTCCTCGAGCGGCGGGGATGTATC
TCCGCCAGGACTAAAGACCGTCACACGAAGGTTAACGGACGAAGCCGTCGAGTTACGATG
CCGGCTCTCGCCGCCGCTAGGATTTTCCAGTTAACCGGTGAGCTCGGTCAAAAACCTGAA
GGAGAAACCATCGAATGGCTTCTTAGTCAAGCTGAACCGTCGATTATTGCCGCCACTGGC
TACGGGACTAAGCTCATTTTCGAATTGGGTTGATGTTGCGGCGGACGATTCTCTCGTCGTCG
TCGTCGATGACGTGCGCGCAAACGCAAACGCAAACGCCACAATCGCCGAGTTGTAGGTTG
GATCTTTGTGACCAATCGGAATTCAGTATCCGGTGAATGGTTACAGTCATATGCCGTTTC
ACAGCGATGCTTTTAGAGCCGATGACCACGACGGCGGAATCTGAGGTTGAGATCGCGGAG
GAGGAGGAACGTAGACGCCGTCACCATTAGTAAAATTAGGCTTTTGATTTAGAGTGTTAA
AATTAGGATTTTAAAGTTTAGGAGGTAACAGATAAGGATAATT
>G1663 Amino Acid Sequence (domain in AA coordinates: TBD)
MIFQNVCRNESNFNAIASESRSQTFGVSKSSSSGGGCSISARTKDRHTKVNGRSRRVTMP
ALAAARIFQLTRELGHKTEGIEWLLSQAEPSIIAATGYGTKLISNVWDVAADDSSSSS
SMTSPQTQTPTQSPSCRDLCPQIGIYQPVNGYSHMPFTAMLLPEMTTAESEVEIAEE
EERRRRHH*
>G1231 (103..870)
CAAACCCAAATCTCTCAGCGCCGGTCAAATACTTGTCTCTCTCTCTCTCTCTTTTAC
TCTTGTCTGTCTCCTTCGAAGCTGTTTGTCTGTGAAGAAAGATGGAAGCAGGTGGCGCG
TACAATCCACGCACTGTTGAAGAGGTGTTTAGGGATTTTAAAGGTCGTAGAGCTGGCATG
ATTAAGGCTTTAACCCTGATGTTTCCAGAGTTTTCGACTTTGTGATCCCGAAAAGGAG
AACCTTTGCTTTTACCGACATCCAAATGAGCACTGGGAAGTGAATTTGCCAGCTGAAGAG
GTTCTCTCTGAGCTCCAGAGCCTGTCTTGGGTATCAATTTGCCAGAGACGGGATGGCG
GAAAAGGATTGGTTGTCCCTTGTGTGCTGCCACAGTATGCTTGGCTTCTTGCTGTGCT
TTCTTTTTTGGAGCCAGGTTTGGATTTGACAAAGCTGATAGGAAGAGGCTTTTCAATATG
GTGAATGACCTCCCAACAATCTTTGAGGTTGTAGCTGGCACTGCTAAGAAACAAGGAAAA
GATAAGTCTCTGTTTCCAACAACAGCAGCAACAGATCAAATCAAGCTCCAAGCGAGGA
TCTGAATCCCGTGCCAAGTTCTCAAAGCCGGAGCCCAAGATGATGAGGAGGAGGAAGAG
GAAGGTGTGGAAGAGGAGGATGAGGATGAGCAAGGTGAAACACAGTGTGGAGCATGTGGT
GAGAGCTATGCAGCTGATGAGTTCTGGATTTGCTGTGACCTCTGTGAGATGTGGTTTCAT
GGAAAGTGTGTTAAGATAACACCAGCAAGAGCTGAGCACATCAAGCAATACAAGTGCCCT
TCTTGACAGCAACAAAAGGGCTCGTTCTTAAATTTGTTGACCGCTCGTTCTGTGTATCTA
CCTTTGCATATGATGATGAACAGCTTAAGTGTGTTGTTAGATCAGATTTGTATATGGA
TTTGGTAATTTAGGAAGACATTTTAGTTTTTTTCATTGTTACATTTTGGCGATTGAAGGGA
TAACTCTTTGTTTAGGGGTAATGATCTTTTGCTCTGTTTATGTTTGTATTAAACATTC
TTCAAACCTCAATCAAAGTATTTTGGTTAGTCTTAAAA
>G1231 Amino Acid Sequence (domain in AA coordinates: TBD)
MEAGGAYNPRTVBEVFRDFKGRRRAGMIKALTTDVQEFFRLCDPEKENLCLYGHNPNEHWEV
NLPAAEVVPELPEPVLGINFARDGMAEKDWLSLVAVHSDAWLLAVAFFGARFGFDKADR
KRLFMVNDLPTIFEVVAGTAKKQKDKSSSVSNSSNRKSSSKRGSESRAKFSKPEPKD
DEEEEEEGVEEEDDEQGETQCGACGESYAADEFWICCDLCEMWFHGKCVKITPARAEHI

KQYKCPSCSNKRARS*

>G227 (21..983)

GTACCGTCGACGATCCGGCGATGTCAAACCCGACCCGTAAGAATATGGAGAGGATTAAAG
GTCCATGGAGTCCAGAAGAAGATGATCTGTTGCAGAGGCTTGTTTCAGAAACATGGTCCGA
GGAAGTGGTCTTTGATTAGCAAATCAATCCCTGGACGTTCCGGCAAATCTTGTCGTCTCC
GGTGGTGTAAACAGCTATCTCCGGAGGTAGAGCACCGTGCTTTTTCGCAGGAAGAAGACG
AGACGATTATTCGAGCTCACGCTCGGTTTGGTAACAAGTGGGCTACGATCTCTCGTCTTC
TCAATGGACGAACCGATAACGCTATCAAGAATCATTGGAACCTCGACGCTGAAGCGAAAAT
GCAGCGTCAAGGGCAAAGTTGTGATTTTGGTGGTAATGGAGGGTATGATGGTAATTTAG
GAGAAGAGCAACCGTTGAAACGTACGGCGAGTGGTGGTGGTGTCTCGACTGGCTTGT
ATATGAGTCCCGGAAGTCCATCGGGATCTGACGTCAGCGAGCAATCTAGTGGTGGTGCAC
ACGTGTTTAAACCAACGTTAGATCTGAGGTTACAGCGTCATCGTCTGGTGAAGATCCTC
CAACTTATCTTAGTTTGTCTCTTCTTGGACTGACGAGACGGTTCGAGTCAACGAGCCGG
TTCAACTTAACCAGAATACGGTTATGGACGGTGGTTATACGGCGGAGCTGTTTCCGGTTA
GAAAGGAAGAGCAAGTGGAGTAGAAGAAGAAGCGAAGGGGATATCTGGTGGATTTCG
GTGGTGGATTTCATGACGGTGGTTTCAGGAGATGATAAGGACGGAGGTGAGGAGTTACATGG
CGGATTTACAGCGAGGAAACGTCGGTGGTAGTAGTTCTGGCGGCGGAGGTGGCGGTTTCGT
GTATGCCACAAAGTGTAAACAGCCGTCGTGTTGGGTTTAGAGAGTTTATAGTGAACCAA
TCGGAATTGGGAAGATGGAGTAGGCGGCC

>G227 Amino Acid Sequence (domain in AA coordinates: 13-112)

MSNPTRKNMERIKGPWSPEEDLLQRLVQKHGPRNWSLISKSIIPGRSGKSCRLRWCNQLS
PEVEHRAFSQEEDETIIRAHARFGNKWATISRLNLGRDINAIKNHWNSTLKRKCSVEGQS
CDFGGNGGYDGNLGEELKRTASGGGGVSTGLYMSPGSPSGSDVSEQSSGGAHVFKPTV
RSEVTASSSGEDPPTYLSLSLPWTDVTVRVNBPVQLNQNTVMDGGYTAEFPVRKEEQVE
VEEEEAKGISGGFGGEFMTVVQEMIRTEVRSYMADLQRGNVGGSSSGGGGGGSCMPQSVN
SRRVGFREFIVNQIGIGKME*

>G1842 (219..809)

ACTATTACATGCCTCTTCCTCGCTTCAAACCGGCACCGTTTCCACTTGTATTATTTTTC
TCTCTATCGTCTAACAAAAAAGCTGACTTGGGATTTTTTTTCAATTGTCTAGCCCA
AAAGAAGAAGATAGAAACGAAGAAAAAAGCAAACACATTTTGGGTCCCCGGTGGTTAGG
ATCAAATTAGGGCACAAACCTTATCGGAGAAAGAAGCCATGGGAAGAAGAAAAGTCGAGA
TCAAGCGAATCGAGAACAAAGCAGTCGACAAGTCACTTTCTCAAACGACGCAAAGGTC
TCATCGAAAAAGCTCGACAACTTCAATTCTCTGTGAATCTTCCATCGCTGTTGTCGCCG
TCTCCGTTCCCGGAAAACTCTACGACTCTGCCTCCGGTGACAACATGTCAAAGATCATTG
ATCGTTATGAAATACATCATGCTGATGAACCTTAAAGCCTTAGATCTTGCAGAAAAAATTC
GGAATTATCTTCCACACAAGGAGTTACTAGAAATAGTCAAAGCAAGCTTGAAGAATCAA
ATGTCGATAATGTAAGTGTAGATTCTCTAATATCTATGGAGGAACAGCTCGAGACTGCTC
TGTCACTAATTAGAGCTAAGAAGACAGAACTAATGATGGAGGATATGAAGTCACTTCAAG
AAAGGGAGAAGTTGCTGATAGAAGAGAACCAGATTCTGGCTAGCCAGGTGGGGAAGAAGA
CGTTTCTGGTTATAGAAGGTGACAGAGGAATGTACGGGAAAATGGCTCCGGCAACAAAG
TACCGGAGACTCTTTCGCTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCACCAT
AAACTTACTCACAGCTGATTCAGAAGCTTTTACAAAATTGTAAATTATAAAAAGCTGCA
TAATAATCTCAACCTTTTTATCTTCTCGCGCCAAATGTGGAAATAAAGGTAAAAACAAAC
GAAGCTCTTTCTTTTATGCGAAAGAATTGTAAACTAAGATAAAGCTACCGATCTTTGT
TGTACCTTAGTAGACAAATATCAGAGTTCTTGTGCTTGT

>G1842 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKVEIKRIENKSSRQVTFSKRRKGLIEKARQLSILCESSIAVVAVSGSGKLYDSASG
DNMSKIIDRYEIHHADELKALDLAEKIRNYLPHKELLEIVQSKLEESNVDNVSVDLSISM
EEQLETALSVIRAKKTELMEDMKSLQEREKLLIENQILASQVGKKTFLVIEGDRGMSR
ENGSGNKVPETLSLLK*

>G1505 (1..681)

ATGGATGATATAGCGGAACCTGAATGGTTATCAAATTCGTAGATGATTCTTCTTTCACG
CCGTATTCTGCTCCGACGAATAAACCGGTTTGGTTAACCGGAAATCGGAGACATCTTGTA
CAACCGGTTAAAGAGGAGACCTGCTTCAAATCCCAACATCCGGCCGTCAAACCCAGACCC
AAACGAGCCAGAACCGGAGTCAGAGTCTGGTCTCATGGTTCCGAGTCGTTAACCGACTCA
TCTTCAAGCTCTACAACATCTTCGTCGCTCTCTCTCGTCTTCAAGCCCTCTATGGCTC
GCCAGCGGTCAAGTTCTTGATGAGCCAATGACTAAACACAAAAGAAGAAGAAAGTTTGG

AAAAACGCTGGTCAGACGCAAACGCAAACGCAGACGCAGACGCGGCAGTGTGGTCATTGT
GGAGTTTCAGAAAACGCCGCAGTGGAGAGCAGGACCATTAGGAGCGAAGACGTTGTGTAAAT
GCGTGTGGTGTGCGTTACAAATCGGGTCGGTTACTACCCGAATATAGACCCGCTTGTAGC
CCAACATTTTCGAGTGAGCTTCACTCAAACCACCACAGTAAAGTCATTGAGATGCGTAGG
AAGAAAGAGACTTCTGACGGTGCTGAAGAAACCGGTTTGAACCAGCCGGTTCAGACGGTT
CAGGTTGTCTCGAGTTTTTGA

>G1505 Amino Acid Sequence (domain in AA coordinates: TBD)
MDDIAELEWLSNFVDDSSFTPYSAPTNKPVWLGNRRHLVQPVKEETCFKSQHPAVKTRP
KRARTGVRVWSHGSQSLTDSSTSSSTSSSSPRSSPLWLASGQFLDEPMTKTQKKKVV
KNAGQTQTQTQTRQCGHCGVQKTPQWRAGPLGAKTLCNACGVRYKSGRLLPEYRPACS
PTFSSELHNSNHHSKVIEMRRKKETSDGAEETGLNQPVQTVQVVSSF*

>G657 (1..2331)
ATGAAGCGTGAGATGAAAGCACCTACTACTCCACTAGAGAGTCTCCAAGGTGACCTCAAA
GGAAAACAAGGAGGACATCTGGCCCTGCTAGACGATCTACCAAAGGACAATGGACACCT
GAAGAGGACGAAGTCTTGTGTAAAGCTGTTGAGCGTTTTCAAGGAAAGAAGTGGAAAG
ATAGCTGAATGTTTAAAGGATCGGACTGATGTTCAAGTGTCTTCATAGATGGCAAAGGTC
TTGAACCCAGAGCTTGTGAAAGGACCGTGGTCAAAGAGGAGGATAACACAATAATTGAC
CTGGTTGAAAAATATGGGCCAAAGAAATGGTCTACTATATCTCAGCATTACCTGGGCGC
ATAGGAAAGCAATGTAGGGAAAGGTGGCATAACCATCTTAACCTGGGATTAATAAAAAAT
GCATGGACTCAGGAAGAGGAACTGACTCTTATTCGTGCGCATCAAATTTATGGGAATAAA
TGGGCAGAGCTTATGAAATTTTGGCCAGGAAGGTGAGATAATTCGATAAAAAATCATTGG
AACAGCTCAGTTAAGAAGAAGTTGGATTCTTACTATGCATCAGGTCTTTTAGATCAGTGT
CAAAGCTCGCCATTAAATTTGCCCTTCAGAACAATCTATCGCTTCATCTTCTCGTGGATG
CACAGCAATGGAGATGAAGGTAGTTCAAGGCCAGGGGTTGATGCTGAGGAATCAGAAATGC
AGCCAAGCTTCAACTGTTTTCTCACAATCAACCAACGATTTACAAGATGAAGTTCAACGT
GGAAATGAGGAATATTACATGCCTGAATTTTCAATCAGGAACGGAGCAGCAAATCTCAAAC
GCTGCATCTCATGCAGAACCGTACTACCTTCTTTAAAGATGTCAAATTTGTTGTCCCC
GAAATTTCTTGTGAAACAGAATGTTTCAAGAAAGTTTCAAGATCTTAATTGTTCTCACGAG
CTAAGAACTACCACAGCTACGGAGGATCAATTGCCGGGTGTATCTAATGATGCTAAACAG
GACCGTGGTCTAGAGTTATTGACCCATAACATGGACAACGGTGGAAAAACCAAGCACTT
CAACAAGATTTTCAAAGTTCAAGTAAGATTAAAGTATCAACCTTTTTTGTCAAACCTCGGAC
ACAGATCCAGAAGCTCAAACTTTGATCACGGATGAGGAGTGTGTAGGGTTCTTTTCCA
GATAACATGAAAGATAGCAGTACATCTTCTGGTGAGCAAGGTGCGGAATATGGTTGACCT
CAAAACGGCAAAGGATCTCTTGTCTCAGGCTGCAGAAACCCATGCTCATGAAACTGGA
AAAGTTCCAGCTTTACCGTGGCATCTTCAAGTTCTGAGGGCCTGGCGGGTCATAATTGT
GTCCCTTTGTTGGATTGAGACTTGAAGGACTCACTTTTACCCCGTAATGATTCCAACGCT
CCTATACAAGGTTGTGCGCTTTTGGAGCTACCGAATTAGAATGTAAGACTGATACAAAT
GACGGTTTCATCGATACTTACGGACATGTAACCTTCCATGGCAATGATGATAATGGTGGT
TTCCCAAGACAACAGGGGCTGTATATATTTCCCAAGGATTCTTTGAAGCTAGTACCTTTG
AATAGTTTTTCTTCTCTCTTAGAGTGAACAAGATTATTTTCTTATTGACGATAAGCCG
GCTGAAAAAGACAAAGGAGCTCTTTGTTATGAACCTCCACGTTTTTCCAAGTGCAGATATT
CCTTTCTTCAGCTGTGATCTTGTACCATCAAATAGTGACTTACGGCAAGAGTACAGTCCC
TTTGGTATCCGTCAGTTGATGATTTCTTCAATGAATTGTACAACCTCCGTAAAGGTTATGG
GATTCACCGTGTACGATAGGAGCCCTGATGTCATGCTTAATGATACTGCCAAAAGTTTT
AGTGGTGCACCATCCATCTTAAAGAAGCGGCATCGAGACTTGCTTTACCTGTGCTTGAT
AGAAGAAAAGACAAAAAGCTTAAAGGGCTGCGACTTCTCTCTTGGCTAATGATTTTTCG
CGTTAGATGTAATGCTTGTGATGAAGGAGATGATTGCATGACCTCTCGTCCGTGAGTCT
CCTGAAGATAAAAAATATGTTGCTTCCATAGCCAGAGATAACAGAAATTGTGCA
TCAGCTCGGTTATATCAAGAAATGATTCCGATAGATGAGGAACCAAGGAAACCTTAGAA
TCAGGTGGAGTGACTTCTATGCAAAATGAAATGGATGTAATGACGGTGGTGTCTCAGCT
AAAAATGTAAGTCCGCTTTTGTCTTGCATATTATCTGGTATCAGTTATAA

>G657 Amino Acid Sequence (domain in AA coordinates: TBD)
MKREMKAPTTPLSLQGLDKGKQGRSTSGPARRSTKGQWTPPEDEVLCKAVERFQKKNWKK
IAECFKDRDQVCLHRWQKVLNPELVKGPWSKEEDNTIIDLVEKYGPKKWSTISQHLPGR
IGKQCRERWHNHLNPGINKNAWTQEEELTLIRAHQIYGNKWAELMKFLPGRSDNSIKNHW
NSSVKKGLDSYYASGLLDQCQSSPLIALQNKSIASSSSWMHNSNGDEGSSRPGVDAEES
SQASTVFSQSTNDLQDEVQRGNEEYMPFHSQTEQQISNAASHAEPYYPSPKDVKIVVP

EISCETECSSKKFQNLNCSHELRTTTATEDQLPGVSNDKQDRGLELLTHNMDNGGKNQAL
 QQDFQSSVRLSDQPFLLNSDTPPEAQTLLITDEECCRVLPFDNMKDSSTSSGEQGRNMVDP
 QNGKGSLSQAAETHAHETGKVPALPWHPPSSSEGLAGHNCVPLLDSDLKDSLPRNDSNA
 PIQGCRLFGATELECKTDTNDGFIDTYGHVTSHGNDNGGFPEQQGLSYIPKDSLKLVLPL
 NSFSSPSRVNKIYFPIDDKPAEKDKGALCYEPPRFPSADIPFFSCDLVPSNSDLRQEYSP
 FGIRQLMISSMNCTTPLRLWDSPOCHDRSPDVMLNDTAKSFSGAPSILKKRHRDLLSPVLD
 RRKDKKLKRAATSSLANDFSRLDVMLDEGDDCMTSRPSESPEKDNICASPSIARDNRNCA
 SARLYQEMIPIDEPEKETLES GGVTSMQENENG CNDGGASAKNVSPSLSLHIIWYQL*

>G1959 (141..1028)

CGTCGACTGTCCATAAATCCGGAGCCTGACCCGACGTTTGACCCGGATCCGAAACTCCCA
 CAATCTCCATACCACCCAAATTCATCTCCCTAAAGCTTTCTCTCACTTTCCCGGAAAA
 TCGGCGACCAAAATTGAAAATGTACTCAGCGATTGCTCGCTTCCACTCGATGGTGGAC
 ACGTTGGTGGTGACTACCATGGACCTCTTGACGGAACCAATCTCCCGGTGACGCTTGTT
 TGGTTTTAACGACTGACCCCTAAACCTCGTCTCCGGTGGACAACCTGAGCTTCATGAGAGAT
 TCGTTGACGCCGTACTCAGCTCGGTGGTCTGACAAAGCGACTCCCAAACTATTATGA
 GAACAATGGGAGTGAAAGGTCTCACTCTTACCACCTCAAATCACATCTTCAGAAATTCC
 GCCTAGGGAGGCAAGCTGGCAAAGAATCAACTGAGAACTCTAAAGATGCTTCTTGTTAG
 GGGAGAGTCAGGACACAGGTTTCTCTCGACATCATCAATGAGAATGGCGCAGCAGGAGC
 AGAACGAGGGTTACCAAGTCACCGAAGCTCTACGTGCTCAGATGGAAGTCCAAAGAAGAC
 TACACGATCAATTGGAGGTGCAACGGAGGCTCCAGCTGAGGATAGAGGCACAAGGAAAT
 ACCTGCAATCGATTCTTGAAAAGCTTGCAAGGCCTTTGACGAGCAAGCTGCTACTTTTG
 CTGGACTTGAGGCTGCTAGGGAAGAGCTATCAGAGCTAGCCATCAAAGTCTCCAATAGCT
 CTCAGGAACATCAGTCCCGTACTTCGATGCAACAAAGATGATGATGATGCCATCGTTGT
 CAGAGCTTGCACTAGCAATAGACAACAAAAACAACATCACAACTGTTTCTAGTAGAAA
 GCTCTCTGACTTCCATCACACATGGGAGCTCTATATCTGCTGCATCAATGAAGAAGCGTC
 AACGTGGAGACAATTTGGGCGTAGGGTATGAATCAGGCTGGATTATGCCTAGTAGACCA
 TTGGATAAAGTTTAGGAGAGGGGAAAAAGTTTATATGGGAAAGGTAGAGATAAGATTAA
 CTGTTCTTTACTTGCTTTGAGGGGCTGCGGCCGCT

>G1959 Amino Acid Sequence (conserved domain in AA coordinates:46-97)

MYSAIRSLPLDGGHVGGDYHGPLDGTNLPDACLVLTTDPKPRLRWTELHERFVDAVTQ
 LGGPDKATPKTIMRTMGVGLTLYHLKSHLQKFRLRQAGKESTENSKDASCVGESQDTG
 SSSTSSMRMAQQEQNEGYQVTEALRAQMEVQRRRLHDQLEVVQRRQLRLIEAQGKYLSILE
 KACKAFDEQAATFAGLEAAREELSELAIKVSNSSQGTSVPYFDATKMMMPSLSELAVAI
 DNKNITTNCSVESSLTSITHGSSISAASMKRQRQDNLGVGYESGWIMPSSTIG*

>G2180 (1..1440)

ATGGCTCCTGTCTCGTTACCTCCAGGTTTCCGATTCCATCCAACAGACGAGGAACATAATT
 ACTTACTATCTAAAAAGAAAGATCAACGGTCTAGAAATCGAACTTGAAGTTATCGCTGAA
 GTTGATCTTTACAAGTGTGAGCCATGGGACTTACCAGGGAAGTCCTTGCTTCCGAGCAAA
 GACCAAGAATGGTACTTCTTCAGCCCACGAGACCGGAAGTATCCAACGGCTCAAGGACA
 AACCGGGCAACTAAAGGCGGTTATTGGAAGGCTACAGGTAAAGACCGCCGAGTTAGTTGG
 AGAGACCGAGCCATAGGAACCAAGAAGACATTGGTTTACTACCGTGGGCGCGGCCACAT
 GGCATAAGAAGTGGTTGGGTATGCACGAATATCGACTTGATGAAACAGAATGTGAGCCT
 TCTGCATACGGCATGCAGGACGCATATGCACCTTTGTCGTGTGTTCAAAAAGATTGTTATT
 GAAGCTAAGCCAAGAGATCAACATCGGTATATGTCCACGCGATGTGCAATGTGAGTGGT
 AATTGCTCATCGAGTTTTGACACTTGTTCGGATCTCGAAATCAGTTCAACTACTCATCAA
 GTTCAAAACACATTCCAACCGCGATTGGCAACGAGCGGATTTAACTCCAACGCAATCAGC
 AACGAGGATTGGTCACAATACTACGGTCTCTTCTTATAGACCGTTCCCTACTCCATATAAG
 GTTAACACAGAGATGGAATGTTCAATGTTACAACACAATATATATCTACCACCGTTGCGT
 GTAGAGAACTCTGCGTTTAGTGATTCCGATTTCTTCACGAGTATGACTCACAACAACGAC
 CATGGCGTTTTCGATGACTTTTACTTTTGGCTGCAAGTAACTCCAACCACAATAATAGCGTT
 GGTGATCAAGTGATCCACGTTGGCAATTATGATGAACAATTAATAACATCTAACCGTCAT
 ATGAACCAGACTGGTTATATAAAAGAGCAGAAGATCAGATCGAGTTTGGATAATACTGAC
 GAAGATCCAGGATTTTCATGGTAACAATACCAATGACAACATAGATATCGATGATTTTCTC
 TCGTTTGATATATATAACGAGGACAACGTGAATCAAATAGAAGATAATGAAGACGTGAAT
 ACAATGAAACCCCTTGATTTCATCGGGATTTCGAGGTGGTTGAAGAAGAACTAGATTTAAC
 AACCAATGCTCATCTCGACATATCAAACGACAAAGATTCTATATCACCAGTCGTACCT
 TGTACACGTTGAAAGTTCAGTCAATCCTATTAGTCACAATGTGGAAGAGAGAACATTG

TTCATTGAAGAGGACAAAGATTCTTGGTTACAAAGAGCTGAGAAGATCACGAAGACAAAA
 CTAACACTTTTTAGTTTAAATGGCTCAGCAATACTACAAATGTCTTGCTATTTTTTTCTGA
 >G2180 Amino Acid Sequence (conserved domain in AA coordinates:7-156)
 MAPVSLPPGFRFHPTDEELITYYLKRKINGLEIELEVIAEVDLYKCEPWDLPKSLPSK
 DQEWYFFSPRDRKYPNGSRITNRATKGGYWKATGKDRRVSWRDRAIGTKKTLVYRGRAPH
 GIRTGWMHEYRLDETECEPSAYGMQDAYALCRVFKKIVIEAKPRDQHRSYVHAMS NVSG
 NCSSSFDTCSLEISSTTHQVQNTFQPRFGNERFNSNAISNEDWSQYYGSSYRPFPTPYK
 VNTEIECSMLQHNIYLPPLRVENSAPSDSDFFTSMTHNNDHGVFDDFTFAASNSNHNSV
 GDQVIHVGNYDEQLITSNRHMNTGYIKEQKIRSSLDNTDEDPGFHGNNTNDNIDIDDFL
 SFDIYNEDNVNQIEDNEDVNTNETLDSSGFVVEEETRFNNQMLISTYQTTKILYHQVVP
 CHTLVHVNPI SHNVEERTLFIEEDKDSWLQRAEKITKTKLTLFSLMAQQYYKCLAIFF*
 >G1817 (1..1308)

ATGAAGGACGCAGAGAAGCGAGAGGTGATTGCATCATCATCATTACAAAGAAAGAGAAAC
 AGAGGAAGAAGACTAAGGAAAAAGAAAGAAACGAGAAGCGAGTACTAATGGTTCCA
 TCATCATTACCAAACGACGTGCTAGAGGAGATCTTTTAAAGATTTCCGGTTAAAGCCCTA
 ATCCGACTCAAGTCTCTCTCGAAACAATGGAGATCGACGATCGAATCTCGCAGTTTTGAA
 GAGAGACACTTGACGATCGCTAAGAAAGCCTTCGTGGATCATCCCAAGGTCATGCTCGTA
 GGAGAAGAAGATCCCATAAGAGGAACCGGGATTCTGTCAGACACTGACATTGGTTTTAGG
 TTATTCTGCTTGGAATCGGCTTCTCTTCTATCCTTTACTCGTCTCAATTTCCCTCAAGGG
 TTCTTCAACTGGATCTACATATCTGAAAGCTGTGATGGCCTTTTCTGCATCCATTCCCCA
 AAATCACATTCCGTATATGTAGTGAATCCGGCTACACGGTGGCTCCGCCACTTCTCTCCG
 GCAGGGTTTCAGATTTTGATCCACAAGTTTAACCCCACTGAACGTGAGTGAATGTAGTG
 ATGAAATCAATCTTTTCATCTAGCATTCTGTGAAGGCCACCGATTACAAATAGTGTGGTTG
 TACAATTGTGATAAGTACATTGTTGATGCGTCGAGTCCAAACGTGGGAGTCACAAAGTGC
 GAGATTTTGACTTTAGGAAAAATGCTTGAGGTTACTTGGCTTGCACTCCAAGTCATCAG
 ATATTCTATTACCAAAGCCAGCATCTGCAAACGGGTCGGTTTTATTGGTTTACAGAACCA
 TATAATGAAAGAATCGAAGTAGTGGCTTTTGATATTCAGACCGAAACATTCCGGTTGCTG
 CCTAAGATTAATCCGGCTATTGCTGGTTTCAGATCCTCACCATATTGACATGTGCACTCTG
 GATAATAGTTTGTGTATGTGCGAAAAGGGAGAAAGATACTATGATCCAAGATATTTGGAGG
 TTGAAACCATCAGAAGACACATGGGAAAAGATTTTAGCATAGACTTGGTTTCTGTCTCT
 TCTTCTCGGACTGAGAAGCGTGATCAATTTGATTGGAGCAAGAAGGATAGGGTTGAGCCA
 GCCACACCCGTCGCGGTTTGTAAAGAATAAGAAGATCCTTCTCTCACATCGCTATTCCCGA
 GGTTTGGTAAAGTACGATCCCCTAACAAATCTATCGATTTTTTTTTCCGGACATCCTACC
 GCTTACAGAAAAGTTATTTATTTTCAAAGTTTGATATCTCATCTATAA

>G1817 Amino Acid Sequence (conserved domain in AA coordinates:47-331)
 MKDAEKREVIASSSLQKRNRGRRLRKRNRNEKRVLMVPSSLPNDVLEEIFLRFPVKAL
 IRLKSLSKQWRSTIESRSFEERHLTIAKKAFVDHPKVMVLVGEEDPIRGTGIRPDIDIGFR
 LFCLESASLLSFTRLNFPQGFNWIYISESCDGLFCIHSPKSHSVYVNPATRWLRLLPP
 AGFQILIHKNFPTERENVMKSIHFLAFVKATDYKLVLVLYNCDKYIVDASSPNVGVTKC
 EIFDFRKNWRYLACTPSHQIFYQKPPASANGSVYWFTEPYNERIEVVAFDIQTETFRLL
 PKINPAIAGSDPHHIDMCTLDNSLCMSKREKDTMIQDIWRLKPS EDTWEKIFSIDLVSCP
 SSRTEKRDQFDWSKKDRVEPATPVAVCKNKKILLSHRYSRGLVKYDPLTKSIDFFSGHPT
 AYRKVIYFQSLISHL*

>G1649 (61..1311)
 ATTCACAAAACCGGAAAAAAGACAAGTAAAGAAAGCTTTGTTTCAGTTTACTTCA
 ATGGAAGCAAAACCTTAGCATCATCATCTGAACCAACATGATTTCTCCATCATCA
 AACATTAACCAAAATTAAGATGAAGATTATATGGAGCTGGTGTGTGAAAATGGGCAG
 ATTCTTGCAAGATTCGAAGACCAAGAACAACGGTTCTTTTCAAAGCAACGTAGGCAA
 TCTCTCCTGGATTTGTATGAGACCGAGTACAGCGAGGGTTTCAAGAAAAACATCAAGATT
 CTTGGAGACACACAAGTTGTTCCGGTGAGTCAGTCTAAGCCACAACAAGATAAAGAAACC
 AATGAACAAATGAACACAATAAGAAGAAGCTAAAGTCTCCAAAATCGAATTTGAGAGA
 AATGTTTCGAAAAGCAACAAATGTGTTGAATCATCAACATTAATTGATGTTTCTGCTAAA
 GGTCCAAAGAATGTTGAAGTTACTACAGCTCCTCCTGATGAGCAATCTGCAGCTGTTGGT
 AGATCCACGGAATTGTATTTGCTTCTCATCGAAGTTTTCTCGAGGAACCTCGAGAGAT
 CTAAGTTGTTGTTCTTTAAAGAGGAAGTATGGAGATATTGAAGAAGAAGATCAACCTAT
 TTAAGTAATAATTCAGATGATGAATCAGATGATGCGAAGACACAAGTTCATGCGAGAAC
 AGAAAGCCGGTGACTAAAAGAAAACGAAGCACAGAAGTCCATAAGTTATATGAAAGAAAA

CGAAGAGATGAATTCAACAAGAAAAATGCGTGCTTTGCAGGACCTACTACCAAATTGTTAC
AAGGATGATAAGGCTTCATTGTTGGATGAGGCTATCAAATATATGCGGACCCCTCAACTT
CAAGTTCAGATGATGAGTATGGGAAATGGATTAATAAGACCACCTACGATGTTGCCAATG
GGTCATTACTCTCCCATGGGTCTAGGAATGCATATGGGTGCAGCAGCAACACCAACATCA
ATACCGCAATTCTGCTATGAATGTTCAAGCAACCGGTTTCCGGGGATGAACAAATGCA
CCACCACAAATGCTAAGCTTTCTTAATCACCCAAGTGGACTAATTCCAAACACTCCTATC
TTTTCTCCATTGGAAAATTGCTCTCAGCCATTCTGTTGCTTCTGTTGTTTCTCAGACT
CAGGCTACTTCTTTTACTCAATCCCAAAGTCTGCGTCCGCTCAAACCTAGAAGATGCA
ATGCAATATAGAGGAAGCAACGGTTTTAGTTATTATCGCTCGCCAACTAATGATTGTA
GAAAGTTGATGTTTCTCCAACCTAACTTTAAGCAAAAAAATGATCGTCTACTCT
GTGTTGTTAGTCTATGGGCTTTTGGGCTTGATTCTTGAACGATTTGAACCTAATTCCA
ACTATTTTCAAAGTGGATGTACAAAGTAAAA

>G1649 Amino Acid Sequence (conserved domain in AA coordinates:225-295)

MEAKPLASSSEPNMISPSSNIKPKLKDEYDYMELVCENGQILAKIRRPKNNGSFQKQRRQ
SLLDLYETEYSEGFKKNIKILGDTQVVPVSQSKPQDKETNEQMNNNKKKLKSSKIEFER
NVSKSNKCVESSTLIDVSAKGPKNVEVTTAPPDEQSAAVGRSTELYFASSSKFSRGTSRD
LSCCSLKRKYGDIEEESTYLSNNSDDESDDAKTQVHARTRKPVTKRKRSTEVHKLKERK
RRDEFNKKMRALQDLLPNCYKDDKASLLDEAIKYMRTLQQLVQMMSMGNGLIRPPTMLPM
GHYSMPGLMHMGAAATPTSIPQFLPMNVQATGFGPMNAPPQMLSFLNHPGLIPNTPI
FSPLENCSPFVVPSCVSQTQATSFTQFPKSASASNLDEAMQYRGSNGFSYRSPN*

>G2131 (69..1010)

GTCTCTCATTTTCATAATTCATTTTCAGGATTGTCTCTCAATCTTTTATTCTTCTCATT
CACCGGTAAATGGCAAAAGTCTCTGGGAGGAGCAAGAAAACAATCGTTGACGATGAAATCA
GCCATAAAACAGCGTCTGCGTCTGAGTCTGCGTCCATTGCCTTAACATCCAAACGCAAC
GTAAGTCGCGCGCTCGAAACGCTCCTCTTCAACGCAGCTCCCCTTACAGAGGCGTCACAA
GGCATAGATGGACTGGGAGATACGAAGCGCATTTGTGGGATAAGAACAGCTGGAACGATA
CACAGACCAAGAAAGGACGTCAAGTTTATCTAGGGGCTTACGACGAAGAAGAGCAGCAG
CAGTGCCTACGACTTAGCAGCATTGAAGTACTGGGACGAGACACACTCTTGAACCTCC
CTTTGCCGAGTTATGACGAAGACGTCAAAGAAATGGAAGGCCAATCCAAGGAAGAGTATA
TTGGATCATTGAGAAGAAAAGTAGTGGATTTTCTCGCGGTGTATCAAAATACAGAGGCG
TTGCAAGGCATCACCATAATGGGAGATGGGAAGCTAGAATTGGAAGGGTGTGTTGGTAATA
AATATCTATATCTTGGAAACATACGCCACGCAAGAAGCAGCAATCGCCTACGACATCG
CGCAATAGAGTACCGTGGACTTAACGCCGTACCAATTTGACGCTCAGCCGTTATCTAA
ACCCTAACGCCGCCCGCGGATAAAGCCGATTCCGATTCTAAGCCCATTGGAAGCCCTAGTC
GCGAGCCCGAATCGTCGGATGATAACAAATCTCCGAAATCAGAGGAAGTAATCGAACCAT
CTACATCGCCGGAAGTGATTCCAATCGCCGGAGCTTCCCGACGATATCCAGACGTATT
TTGGGTGTCAAGATTCCGGCAAGTTAGCGACTGAGGAAGACGTAATATTGATTGTTTCA
ATTCTTATATAAATCCTGGCTTCTATAACGAGTTTGATTATGGACCTTAATCGTATTTTC
TACAAGTTTTGTTTGGATTATCTACACAATACATCAATATATTCT

>G2131 Amino Acid Sequence (conserved domain in AA coordinates:50-186, 112-183)

MAKVSGRSKTIVDEISDKTASASESASIALTSKRKRKSPPRNAPLQRSSPYRGVTRHR
WTGRYEAHLWDKNWNTQTQKGRQVYLGADEEEAARAYDLAALKYWGRDTLNFPPLP
SYDEDVKEMEGQSKEEYIGSLRRKSSGFSRGVSKYRGVARHHHNGRWEARIGRVFGNKYL
YLGTYATQEEAAIAYDIAAIEYRGLNAVTFNFDVSRYLNPNAADKADSDSKPIRSPSREP
ESSDDNKSPKSEEVIEPSTSPVIPTRRSFPDDIQTYFGCQDSGKLATEEDVIFDCFNSY
INPGFYNEFDYGP*

>G215 (1..1110)

ATGACTCGTCGGTGTTCGCATTGTAGCAACAATGGGCACAATTCACGCACGTGTCCAACG
CGTGGGTCTGGTTCTCTCCGCCGTGAAGTTATTGGTGTGAGGTTAACGGATGGCTCG
ATTATTAAGAGAGTGCAGATATGGGTAACTCTCTCGGCATTGGCTGTGTGCGGCGCGCG
GCAACGCACCACCGTTTATCTCCGTCTCTCTCTGGCGACGTCAAATCTTAATGATTGCG
CCGTTATCGGATCATGCCGATACTCTAATTTGCATCATAATGAAGGGTATTTATCTGAT
GATCTGCTCATGGTCTGGGTCTAGTCACCGTCGTGGTGAGAGGAAGAGAGGTGTTCTCT
TGGACTGAAGAGGAACATAGACTATTCTTAGTCGGTCTTCAGAACTCGGGAAAGGAGAT
TGGCGCGGTATTTTCGAGAACTATGTAACGTCAAGAACTCTACACAAGTGGCTAGTCAT
GCTCAAAAGTATTTTATTCGACATACTAGTTCAAGCCGCAGGAAAAGACGGTCTAGCCTC
TTCGACATGGTTACAGATGAGATGGTAACCGATTATCGCCAACACAGGAAGAGCAGACC

TTAAACGGTTCTCTCCAAGCAAGGAACCTGAAAAGAAAAGCTACCTTCCTTCACTTGAG
CTCTCACTCAATAATACCACAGAAGCTGAAGAGGTGCTAGCCACGGCGCCACGACAGGAA
AAATCTCAAGAAGCTATAGAACCATCAAATGGTGTTCACCAATGCTAGTCCCGGGTGGC
TTCTTTCCTCCTTGTGTTTTCCAGTGACTTACACGATTTGGCTCCCTGCGTCACTTACGGA
ACAGAACATGCCTTAAACGCTGAGACTTCTTCTCAGCAGCATCAGGTCCCTAAACCAAAA
CCTGGATTTGCTAAAGAACGTGTGAACATGGACGAGTTGGTCGGTATGTCTCAGCTTAGC
ATAGGAATGGCGACAAGACACGAAACCGAAACTTCCCTTCCCGCTATCTTTGAGACTA
GAGCCCTCAAGGCCATCAGCGTTTCACTCGAATGGCTCGGTAAATGGTGCAGATTTGAGT
AAAGGCAACAGCGCGATTCAGGCTATCTAA

>G215 Amino Acid Sequence (domain in AA coordinates: TBD)

MTRRCSHCSNNGHNSRTCPTRGSGSSSAVKLFGVRLTDGSIKKASAMGNLSALAVAAAA
ATHHRLSPSSPLATSNLNDSPSLDHARYSNLHHNEGYSDDPAHSGSSSHRRGERKRGVP
WTEEHRLFLVGLQLKLGKGDWRGISRNYVTSRTPTQVASHAQYFIRHTSSSRKRSSSL
FDMVTDEMVTDSSTQEEQTLNGSSPSKEPEKKSYPLESLNNTTEAEVATAPRQE
KSQEAIEPSNGVSPMLVPGGFFPPCFPVITYIWLPAHLHGTEHALNAETSSQQHQVLKPK
PGFAKERVNMDELVGMSQLSIGMATRHETETSPSPLSLRLEPSRPSAFHSNGSVNGADLS
KGNIAIQAI*

>G1508 (1..420)

ATGCTAGATCACAGTGAAAAGGTCTTATTGGTTGATTAGAAAACCATGAAAACAAGAGCT
GAAGATATGATCGAACAGAACACACTAGTGTAAACGACAAGAAGAAGACTTGTGCTGAT
TGTGGAACAGTAAACTCCTCTTTGGCGTGGTGGTCTGTGGTCCAAAGTCGTTGTGT
AACCGGTGTGGGATCAGAAACAGAAAGAGAAGAGGAGGAACAGAGATAATAAGAAA
TTAAAGAAATCGAGTTCTGGCGGCGGAAACCGTAAATTTGGTGAATCGTTAAACAGAGT
TTGATGGATTTGGGGATAAGGAAGAGATCAACGGTGGAGAAGCAACGACAGAAGCTTGGT
GAAGAAGAACAGCCGTGTGTTACTCATGGCTCTTTCTTATGGCTCTGTTTACGCTTAG

>G1508 Amino Acid Sequence (domain in AA coordinates: 38-63)

MLDHSEKVLVLDSETMKTRAEDMIEQNNTSVNDKKKTCADCGTSKTPLRGGPVGPKSLC
NACGIRNRKRRGGTEDNKKLKKSSSGGNRKFESLQKSLMDLGIKRSTVEKQRQKLG
BEEQAAVLLMALSYGSVYA*

>G2110 (36..1622)

GAGAGCTAATAAAAAATTTATCAAAGAAGACTAATATGGAGAAGGACGATTTCTTGAGGA
GTGGTCATGGAAGAGAAGAAAGCCATGATGAGATGAGAAAACCTTGATTCATCTCAGCATG
ATTCTCATCAAGAACACGACCATATTATAAGATCCAAGTTGGACTCAACTAAAGTCGAAA
TGGATGAGGCTAAAGAGGAAAATCGAAGACTAAAGTCATCATTGAGTAAATCAAGAAAG
ATTTTGACATCCTTCAAACACAATACAACCAATTAATGGCCAAACATAACGAACCAACCA
AGTTCCAATCAAAAGGGCATCATCAAGACAAAGGCGAAGATGAAGACAGAGAAAAAGTTA
ACGAACGTGAAGAACTTGTCTCGTTGAGCCTAGGCAGACGGTTAAATTGAGAGTTCCAA
GTGGTTGGAATAAAGAAAGCAATTAAGATGTTGAAGAAGCGGAAGGTGACAGAAATT
ATGATGATAATGAAAAAAGCAGTATTCAAGGGTTGAGTATGGGGATTGAATACAAGGCTT
TGAGTAATCCTAATGAGAAGTTAGAGATTGATCATAATCAAGAAACCATGTCGTTGGAGA
TTAGTAACAATAAAGATCAGATCACAAAATAGTTTGGGTTAAGAATGATGGAGATG
ATCATGAAGATGAAGATGAGATTTGCCTCAAACCTTGTTAAGAAAACCTAGGGTTTCGG
TGAGATCAAGATGTGAGACACCAACGATGAACGACGGATGTCAATGGAGGAAATATGGCC
AGAAAATAGCTAAAGGCAATCCATGTCCCCGAGCTTACTATCGTTGCACCATGTCAGCTT
CTTGTCAGTAAGAAAACAGGTGCAAAGATGTTGAGAAGATATGTCTATACTTATCTCAA
CGTACGAAGGAACACATAACCATCCACTTCCCATGTCAGCAACTGCCATGGCCTCTGCCA
CTTCCGCTGCCGCTCCATGCTTCTCTCCGGCGCCTCCTCCTCATCCGCCGAGCTG
ATCTTCATGGCCTTAACCTCTCTCTTCCGGCAACAACATCACTCCAAACCTAAAACTC
ATTTCTCCCAATCCCCTTCTTCTTCTGGCCATCCGACCGTCACTCTCGACCTCACAACT
CCTCCTCGTCGAGCAACCGTTCTTATCAATGCTCAATAGATTCAGCTCTCTCCAAGTA
ATGTCTCAGCATCTAATAGTTATCCTTCAACCAATCTCAACTTTTCAAACAACCAACA
CATTGATGAATTGGGGTGGTGGTGAATCCAGTGATCAATACCGTGCAGCTTACGGCA
ACATTAACACCCCATCAGCAATCACCTTACCACAAAATCATTCAAACCCGAACCGCCGGGT
CATCTTTGATCCGTTTGGAGATCATCTTCATCACATTCTCCACAAATAAATCTTGATC
ATATCGGAATCAAGAACATCATCAGTCAACCAAGTGCCATCTTTACCGGCTGAAACAATCA
AGGCAATCACGACAGATCCAAGTTTCCAATCGGCTTTGGCGACAGCTCTATCTTCCATCA
TGGGCGGCGATTAAAGATTGATCACAATGTGACTAGAAATGAAGCTGAGAAGAGCCCTT

AAAGAGAATTGTTATATATATGTTCTTATATACTCAGTACATTGGTAAATGGGTTTAGAC
 TTTCAGTAGTTTCTAGTTTCATCTATATATTGGTTGTTAATCACAAGTTATTTTGTG
 TTGGAGTTTATGGAACATAATGTGTACATATGAACTTTAGAACGAATAAAATAAACTTGG
 AATTCCTTTTAAAAAAAAAAAAAAAAAAAA

>G2110 Amino Acid Sequence (conserved domain in AA coordinates:239-298)

MEKDDFLRSCHGREESHDEMRKLDSSHDDSHQEHDIIRSKLDSTKVEMDEAKEENRRLK
 SSLSKIKKDFDILQTYNQLMAKHNEPTKFQSKGHHQDKGEDEDREKVNREELVSLSLG
 RRLNSEVPSGSNKEEKNKDVEEAEGDRNYDDNEKSSIQGLSMGIEYKALSNPNEKLEIDH
 NQETMSLEISNNKIRSQNSFGFKNDGDDHEDEDEILPQNLVKKTRVSVSRRCETPTMND
 GCQWRKYGQKIAGNCPRAYRCTIAASCPVRKQVQRCESEDMSILISTYEGTHNHLPLM
 SATAMASATSAASMLLSGASSSSSSAAADLHGLNFSLSGNNTPKPKTHFLQSPSSSGHP
 TVTLDLTSSSSSQPFSLMLNRFSSPPSNVSRNSYPTNLNFSNNTNTLMNWGGGNGPS
 DQYRAAYGNINTHQQSPYHKIIQTRTAGSSFDPFGRSSSSHPQINLDHIGIKNIISHQV
 PSLPAETIKAITDPSPFQALATALSSIMGGDLKIDHNVTRNEAEKSP*

>G2442 (71..997)

TCGACCAATTTAGACCATTCCAAATTCGTCGTCCTTTTCTCTGTGTAGTCTAATTATATA
 TTACAAGTAGATGAATTGGTTACCTGAAGCTGAAGCTGAGGAGCACTTGAAAGGTATTCT
 CTCTGGTGATTTCTTTGATGGTCTCACCATCACCCTTGATTGCCCACTTGAGACATCGA
 TTCCACCAATGGTGAGGGAGATTGGGTCGCCAGGTTTCAAGACCTTGAGCCTCCTCCCTT
 GGATATGTTCCCTGCTTTGCCTTCTGACCTCACCTCTTGTCCTCAAGGGCGCCGCTCGTGT
 GCGGATTCCTCAACATGATTCTGCTTGAAGCAGTCTGTTCTTCTGAAGCCTTGTC
 CGGCATTATAGCACTCCCCACCAATCTTCAGCTCCTCTGATATCAAAGTTTCATATCT
 ATTTCACTCTCTAACTCCAGTGTCTAGTTCTCGAGAACAGTTATGGTTCTCTCTCCACCCA
 AAACTCCGGATCTCAGAGATTGGCTTTCCCTGTGAAAGGCATGAGAAGCAAGCGCAGACG
 CCCCACAACAGTGAGACTTAGCTACCTTTTCCCTTTGAACCCAGAAAGTCAACTCCGGG
 TGAATCAGTAACCGAGGGTTACTATTCTTCTGAGCAACATGCCAAGAAGAAGCGCAAGAT
 TCATCTGATCACCCACACCGAGTCTTCCACTTTGGAGTCAAGTAAGTCGGATGGGATAGT
 CCGGATATGCACTCATTGTGAGACAATCACGACCCACAGTGGAGGCAAGGACCCAGTGG
 ACCCAAGACCTCTGCAACGCTTGCAGGAGTCCGGTTCAAATCTGGTCGCCTAGTTCCAGA
 ATACCGGCCAGCCTCAAGCCCGACCTTCATCCCATCTGTGCATTCAAACCTCACACAGGAA
 GATCATTGAGATGAGAAAGAAGGACGACGAGTTTGATACCAGCATGATTTCGAGTGATAT
 CCAGAAGGTAAAGCAGGGGAGGAAGAAATGGTATAAAAGTA

>G2442 Amino Acid Sequence (domain in aa coordinates: 220-246)

MNWLPEAEAEHLKGLSGDFFDGLTNHLDCPLEDIDSTNGEGDWVARFQDLEPPPLDMF
 PALPSDLTSCPKGAARVRIPNMIPALKQSCSSEALSGINSTPHQSSAPPDIKVSYLFQS
 LTPVSVLENSYGLSTQNSGSQRLAFPVKGMRSKRRRPTTVRLSYLFPFEPKSTPGESV
 TEGYYSSEQHAKKKRKHILITHESSTLESSKSDGIVRICTHCETITTPQWRQGPSGPKT
 LCNACGVRFKSGRLVPEYRPASSPTFIPSVHSNSHRKIIEMRKKDDEFDTSMIRSDIQKV
 KQGRKKMV*

>G1051 (66..1031)

CCTGTAAATTAGATTGCTTTCTTTGGTAATCTTTGGATCAAGATCCATCTATTTTTT
 CTTCAATGGCACAACTCCCTCCTAAAATCCCCAACATGACACAACATTGGCCTGATTTCT
 CTTCCCAAAAGCTCTCTCCTTTCTCTACCCCAACCGCAACCGCTGTGCGCCACCGCTACAA
 CCACCGTACAAAACCCCTCATGGGTCGACGAATTCCTCGACTTCTCAGCGTCTCGCCGTG
 GCAACCAACCGTCGTTCCATCAGCGACTCTATCGCATTCCTCGAAGCTCCAACAGTCAGCA
 TCGAAGACCACCAATTGACAGGTTGATGACGAACAGTTTCATGTCGATGTTACCGACG
 ACGACAACCTTCATAGCAATCCTTCCCATATCAACAACAAAATAACAATGTGGGGCCCA
 CGGGATCTTCCGTCGAACACATCCACGCCGTCCAATAGCTTCAACGACGATAACAAAGAAT
 TACCACCGTCCGTCATAACATGAACAATAATATCAACAACAATAACGATGAAGTCC
 AAAGCCAAATGCAAGATGGAGCCAGAAGATGGTACGGCGTCGAATAACAATTCGGGTGATA
 GCTCCGGCAACCGGATTCTCGATCCCAAAAGGGTTAAGAGAATATTAGCAATCGGCAAT
 CAGCACAGAGATCAAGGGTGAGGAACTGCAATACATATCAGAGCTCGAACGTAGCGTCA
 CTTGCTGTCAGGCGGAAGTGTCTAGTGTATCGCCAAGAGTTCATTCTTGATCATCAAC
 GTTTGCTTCTTAACGTTGACAACAGCGCTCTCAAGCAACGAATCGCTGCTTTATCTCAAG
 ACAAGCTTTTCAAAGACGCACATCAAGAAGCATTGAAGAGAGAAATAGAGAGACTTCGAC
 AAGTGTATAATCAACAAAGCCTCACGAATGTGGAAATGCAATCATTATCGGCGACCG
 GAGCCGGTGCTACTCCGGCCGTCGACATCAAGTCGTCCTTGAACAGAGCAGCTCCTCA

ATGTCTCATAAATTAACCATCATGCATCATCAACATTTCTCTCTTTTAGCTTCTTGG
CAAAAGTTCTTGACTATAAAATCTCTTTTCGGGTAAGAAATTCAGGAGATATACATTTTTT
ATTCTAATCACATTGTTTTTAAGTTGTGATGAATTCAGTTTGATGTATCTTATTTATTTT
GTTTATGTCGTCTTTTTTTCTTGGGGTTGATGGAAGGGAATCATCAATTGTTGTTGTAC
AAAGAACTAGTTGAATTTTTTTTTTTTTTTT

>G1051 Amino Acid Sequence (domain in AA coordinates 189-250)
MAQLPPKIPNMTQHWPDPFSSQKLSPFSTPTATAVATATTTVQNPSWVDFLDFSASRRGN
HRRSISDSIAFLEAPTVSIEDHQFDRFDDEQFMSMFTDDDLHNSNPISHNNKNNNVGPTG
SSSNTSTPSNSFNDDNKELPSPDHNMNNNNNNYNDEVQSQCMEPEDGTASNNNSGDS
GNRILDPKRVKRILANRQSAQRSVRKLOQYISELERSVTSLOAEVSVLSPRVAFLDHQRL
LLNVDNSALKQRIAAALSQDKLFKDAHQEALKREIERLRQVYNQQLTNVENANHLSATGA
GATPAVDIKSSVETEQLLNVS*

>G1052 (138..1127)
TGATCATCTAAACTTTCAATTTCTCTCTTGATCCTCACTTGAATTTTTTGTGTTTCTC
TCAAAATCTTTGATCCTTTCTTTGTTTTTCATTGACCTCTTACAAAAAATCTGGTGTG
CCATTAAATCTTTATTAATGGCACAACTTCTCCGAAAATCCCAACCATGACGACGCCAA
ATTGGCCTGACTTCTCTCCAGAACTCCCTTCCATAGCCGCAACGGCGGCAGCCGCAG
CAACCGCTGGACCTCAACAACAAAACCTTCATGGATGGATGAGTTTCTCGACTTCTCAG
CGACTCGCGTGGGACTACCGTCTGTCTATAAGCGACTCCATTGCTTTCCTTGAACCAC
CTTCTCCGCGCTCGGAAACCACTTCGATAGGTTTGACGACGAGCAATTCATGTCCA
TGTTCAACGACGACGTACACAACAATAACCAATCATCATCATCACAGCATCAACG
GCAATGTGGGTCCACGCGTTCATCTCCAACACCTCCACGCGCTCCGATCATAATAGCC
TTAGCGACGACGACAACAACAAGAAGCACCACCGTCCGATCATGATCATCATGAGACA
ATAATGTAGCCAATCAAAACAACGCGCGCGGTAACAATTACAACGAATCAGACGAGGTCC
AAAGCCAGTGCAAGACGAGCCACAAGATGGTCCGTCGGCGAATCAAACTCCGGTGGAA
GCTCCGGTAATCGTATTCACGACCTAAAAGGGTAAAAGAAATTTAGCAAATAGGCAAT
CAGCAGAGATCAAGGGTGAGGAAATGCAATACATATCAGAGCTTGAAAGGAGCGTTA
CTTCATTGCAGACTGAAGTGTAGTGTATCGCCAAGAGTTGCGTTTTTGGATCATCAGC
GATTGCTTCTCAACGTCGACAATAGTGTATCAAGCAACGAATCGCAGCTTTAGCACAAG
ATAAGATTTTCAAAGACGCTCATCAAGAAGCATTGAAGAGAGAAATAGAGAGACTTCGAC
AAGTATATCATCAACAAAGCCTCAAGAAGATGGAGAATAATGTCTCCGATCAATCTCCGG
CCGATATCAAACCGTCCGTTGAGAAGGAACAGCTCCTCAATGTCTAAAGCTGTTCGTTCA
CTAAGATCTTTCTTTTCATGGCGAAAAGATTCTTGACTATAAAACCTCTTTGTGTCAAGA
AATTAATTTATCAAAGAAGATGGCCTTTTTTATTTGATCTAATCACATTTTTTTAAGTTG
TGATGAATTTGCTTTTGATGTATCTGTTTTTTTTTTTTTTTTTT

>G1052 Amino Acid Sequence (domain in AA coordinates 201-261)
MAQLPPKIPTMTTPNWPDPFSSQKLPSIAATAAAAATAGPQQNPSWMDLDFSATRRGT
HRRSISDSIAFLEPPSSGVGNHFDREDFDEQFMSMFTDNDVHNNNNHHHHHSINGNVGPT
RSSSNTSTPSDHSLSDDNNKEAPPSPDHMDNNVANQNNNAAGNNYNESDEVQSQCKT
EPQDGPANQNSGGSSGNRIHDPKRVKRILANRQSAQRSVRKLOQYISELERSVTSLOTE
VSVLSPRVAFLDHQRLLLNVDSAIKQRIAAALAQDKIPKDAHQEALKREIERLRQVYHQQ
SLKKMENNVSQSPADIKPSVEKEQLLNVS*

>G1079 (1..1995)
ATGGGTTGTGCTGCTTCAAGAATTGATAATGAAGAAAAGGTTTTAGTGTGTAGGCAGAGA
AAGAGGCTAATGA AAAAGTTATTAGGGTTACAGGGGAGAATTTGCAGATGCACAGTTGGCT
TATCTTAGAGCTTTGAGGAACACTGGTGTTACTCTTAGGCAATTCAGTGAGTCTGAGACC
TTGGAGCTTGAAAACACTAGTTATGGTTTAAGTTTGCTTTGCCTCCTTCGCCTCCTCCT
ACATTGCCTCCTTCACCTCCACCACCTCCTCCATTTAGCCCGGATTTGAGAAATCCTGAG
ACTAGTCATGACTTGGCTGATGAGGAGGAAGAGGGTGAAAATGATGGTGGTAATGATGGA
AGTGGTGCAGCTCCTECGCCTCCATTGCCGAATCTTGGAACATTTGGAACCTTTTGAG
TCATTGAGCTGCATAGTCATCCAAATGGTGACAATGTAGTTACACAAGTTGAACTGAAG
AAGAAACAACAAATTCAGCAAGCTGAAGAGGAAGATTGGGCGGAGACGAAGTCTCAATTT
GAGGAAGAAGATGAGCAACAAGAAGCAGGAGGTACTTGCTTGATTTGAGTGTTTCATCAA
ATAGAGGCTGTTAGTGGCTGTAAACATGAAGAAGCCAGTCGTCTGAAGTTTAAGCTGGGA
GAAGTTATGGACGGTAACCTCATCTATGACAAGCTGCTCCGGTAAAGATCTTGAGAAAAC
CATGTGACTGATTGTAGAATCAGGAGGACCTTAGAAGGAATCATCAGAGAGTTGGATGAT
TATTTCTTAAAGCATCGGGTTGCGAGAAGGAGATAGCTGTGATAGTAGACATCAACAGT

AGGGATACTGTTGATCCTTTTCAGGTACCAGGAAACAAGAAGGAAGAGAAGCAGCTCGGCA
 AAGGTATTTCAGTGCATTGTTCATGGAGTTGGTCTTCAAAGTCTCTTCAGTTGGGCAAAGAT
 GCTACAACAAGCGGGACTGTTGAACCTGTAGGCCTGGAGCTCACTGCAGCACACTTGAG
 AAGCTATACACAGCTGAGAAGAACTTTACCAGCTAGTCAGAAACAAGAGATTGCCAAA
 GTGGAGCATGAGAGGAAGTCTGCATTACTGCAAAAGCAAGATGGGGAAACCTATGATTTG
 AGCAAAATGGAGAAAGCACGCTTGTCTTTGGAGAGTTTGGAAACCGAGATACAGCGTCTA
 GAAGATTCCATAACTACAACACGCTCATGTTTGTCTTAACCTTGATCAATGATGAGCTGTAT
 CCGCAGCTAGTTGCTTTAACTTCAGGGCTAGCACAGATGTGGAAAACAATGCTCAAGTGT
 CATCAAGTTCAAATTCATATATCCCAGCAACTGAACCATCTTCCGATTACCCGAGTATA
 GATCTCAGTTTCGGAATACAAACGCCAGGCGGTAAATGAAGTAGAGACCGAGGTTACTTGC
 TGGTACAATAGCTTTTGCAGTTAGTAAATTCACGCGAGAATACGTGAAAACACTCTGT
 ACGTGGATCCAACCTACTGATCGCCTCTCTAACGAAGACAACCAAAGAAGTAGCTTGCCT
 GTTGTCTGCTCGTAAGCTCTGCAAAGAGTGGCAGCTTGAATACAACCTGCGTAGGAAATGC
 AATAAACTTGAGAGGAGGCTTGAGAAAGAGCTAATTTCACTGGCTGAGATTGAAAGAAGG
 CTCGAGGGGATTTTAGCAATGGAAGAGGAGGAAGTAAGCTCAACGAGTTTGGGCTCTAAG
 CATCCGTTGTCAATCAAACAAGCCAAGATCGAAGCCTTGAGAAAACGAGTGGATATTGAG
 AAACTAAGTACTTAACTCGGTTCGAGGTTAGTAAGAGAATGACACTAGACAACCTCAA
 TCAAGCCTTCCAATGTCTTTAGATGTTGACTGCTCTAGCTAATGTCTTTGCCAATGGG
 TTTGAATCCGTTAATGGCCAAACCGGTACAGATGTTTCCGACACATCCCAACATTCCGAT
 GAATCTCAACCCTAA

>G1079 Amino Acid Sequence (conserved domain in AA coordinates:1-50)

MGCAASRIDNEEKVLVCRQRKRLMKLLGFRGEFADAQLAYLRALRNTGVTLRQFTBSET
 LELENTSYGLSLPLPPSPPTLPSPPPPPFSPDLRNPETSHDLADEEEEEENDGGNDG
 SGAAPPPPLPNSWNIWNPFESELELHSHPNGDNVVTQVELKKKQQIQQAEEEDWAETKSQF
 EEDEEQQEAGGTCLDLVSHQIEAVSGCNMCKPRRLKFKLGEVMDGNSSMTSCSGKDLEKT
 HVTDCRIRRTLEGIIRELDDYFLKASGCEKEIAVIVDINSRDTVPFRYQETRRKRSSSA
 KVFSALSWSWSKSLQLGKDATTSGTVEPCRPGAHCSTLEKLYTAEKKLYQLVRNKEIAK
 VEHERKSALLQKQDGETYDLSKMEKARLSLESLETEIQRLSDSITTTSRCLLNLINDELY
 PQLVALTSGLAQMWKMLKCHQVQIHISSQLNHLDPDYSIDLSSEYKRQAVNELETEVTC
 WYNSFCKLVNSQREYVKTLCTWIQLTDRLSNEDNQRSSLPVAARKLCKEWQLEYNLRRKC
 NKLERRLEKELISLAEIERRLEGILAMEEEEEVSSTSLGSKHPLSIKQAKIEALRKRVDIE
 KTKYLNSEVSKRMTLDNLKSSLPNVFQMLTALANVFANGFESVNGQTGTDVSDTSQHS
 ESQP*

>G1335 (56..667)

TTTTTTTTTAAAAGATTTAGAGAGAAAAGTGAGTTATTAAGAGATTCCAATCAAAATGAG
 CGGAGACAACGGCGGTGGTGAGAGCGCAAAGGCTCCGTCAAGTGGTTTGATACCCAGAA
 GGGTTTCGGCTTCATCACTCCTGACGACGGTGGCGACGATCTCTTCGTTCCAGTCCCTC
 CATCAGATCTGAGGGTTTCCGTAGCCTCGCTGCCGAAGAAGCCGTAGAGTTCGAGGTTGA
 CATCGACAACAACAACCGTCCCAAGGCCATCGATGTTTCTGGACCCGACGGCGCTCCCGT
 CCAAGGAAACAGCGGTGGTGGTTTCATCTGGCGGACGCGCGGTTTCGGTGGAGGAAGAGG
 AGGTGGACGCGGATCTGGAGGTGGATACGGCGGTGGCGGTGGTGGATACGGAGGAAGAGG
 AGGTGGTGGTTCGAGGAGGCAGCGACTGCTACAAGTGTGGTGGAGCCCGGTACATGGCGAG
 AGACTGTTCTGAAGGCGGTGGAGGTTACGGAGGAGGCGCGGTGGCTACGGAGGTGGAGG
 CGGATACGGCGGAGGAGGTGGTGGTTACGGAGGTGGTGGCCGTGGAGGTGGTGGCGGCGG
 GGGAAAGCTGCTACAGCTGTGGCGAGTCCGGACATTTCCGACGGGATTGCACCAGCGGTGG
 ACGTTAAAACCAACGCCGTTACGCGGTGGAGAAGAGTGAGTTGGTTATCTCACAAGTGA
 TCGGTTCTTTCTCCCGCCGCTTCTATCTCTATTATCCACTTTTGTCTATTATGATG
 GATCTCTATCTTTGTAGTTGGTTTTTCTTGATGGTTTCGATTAGGACTCTTCTTTTG
 GTTTTGCTACTTATGGTTGGTTTTTATTATGGTACTTGTGATATGGGTGAAATGCTCTAC
 TTGTTGCTCTGTTTCAAGTGTTTATAATATGCGAACAATATTCTGGGTTTTGTTTCAA
 AAAAA

>G1335 Amino Acid Sequence (domain in AA coordinates: 24-43, 131-144, 185-203)

MSGDNGGGERRKGSVKWFDTPQKGFITPDDGGDDLFDVHQSSIRSEGFSLAAEEAVEFE
 VEIDNNNRPKAIDVSGPDGAPVQGNSSGGSSGGRRGGFGGGRRGGSGGGYGGGGGGYGG
 RGGGRRGSDCYKCEP GHMARDCEGGGGYGGGGGGYGGGGYGGGGGGYGGGGRRGGGG
 GGGSCYSCGESGHFARDCTSGGR*

>G157 (31..621)

GGGCATAACCCCTTATCGGAGATTTGAAGCCATGGGAAGAAGAAAAATCGAGATCAAGCGA
ATCGAGAACAAAAGCAGTCGACAAGTCACTTTCTCAAACGACGCAATGGTCTCATCGAC
AAAGCTCGACAACCTTTCGATTCTCTGTGAATCCTCCGTCGCTGTTGTGTCGTATCTGCC
TCCGGAAAACTCTATGACTCTTCTCCGGTGACGACATTTCCAAGATCATTGATCGTTAT
GAAATACAACATGCTGATGAACCTTAGAGCCTTAGATCTTGAAGAAAAAATTCAGAATTAT
CTTCCACACAAGGAGTTACTAGAAACAGTCCAAAGCAAGCTTGAAGAACCAAATGTCGAT
AATGTAAGTGTAGATTCTCTAATTTCTCTGGAGGAACAACTTGAGACTGCTCTGTCCGTA
AGTAGAGCTAGGAAGGCAGAACTGATGATGGAGTATATCGAGTCCCTTAAAGAAAAGGAG
AAATTGCTGAGAGAAGAGAACAGGTTCTGGCTAGCCAGATGGGAAAGAATACGTTGCTG
GCAACAGATGATGAGAGAGGAATGTTTCCGGGAAGTAGCTCCGGCAACAAAATACCGGAG
ACTCTCCCGCTGCTCAATTAGCCACCATCATCAACGGCTGAGTTTTACCTTAAACTCAA
AGCCTGATTATCATAATTAGAGAATAAATTTGTATATTATAAAAAGCTGTGTAATCTCAA
CCTTTTATCTTCTCTAGTGTGGAATTTAAGGTCAAAAGAAAACGAGAAAGTATGGATC
AGTGTGTACCTCCTTCGGAGACAAGATCAGAGTTTGTGTGTTGTGTCTGAATGTACGG
ATTGGATTTTTAAAGTTGTGCTTTCTTTCTTCAAAAAAAAAA

>G157 Amino Acid Sequence (domain in AA coordinates: 2-57)

MGRRKIBIKRIENKSSRQVTFSKRRNGLIDKARQLSILCESSVAVVVVSASGKLYDSSSG
DDISKIIDRYEIQHADELRLALDLEEKIQNYLPHKELLETVQSKLEEPNVNDVSVDSLISL
EEQLETALSVSRRKAELMMEYIESLKEKEKLLREENQVLASQMGKNTLLATDDERGMFP
GSSSGNKIPETLPLLN*

>G1895 (1..954)

ATGAATAACCAATCTGTTACTGACAATACAAGTCTTAAGCTGTCTAATCTTAACAAC
GAGTCAAAAAGAAACATCTGAGAACAGTGATGACCAACACAGCGAGATCACAACAATTACA
TCGGAAGAAGAGAAAACAACCTGAAGTGAAGAAACCAGACAAGATTCTTCCATGTCCGAGA
TGCAACAGCGCAGACACCAAATCTGTACTACAACAACCTACAACGTTAACCAGCCACGT
CACTTCTGTAGAAAATGCCAGAGGTATTGGACCGCTGGTGGATCCATGAGGATCGTCCCG
GTTGGCTCAGGCCGTCGCAAGAACAAGGGATGGGTTTCTTCAGACCAGTACCTGCACATC
ACTTCCGAGGATACTGACAATTACAATAGCTCCTCAACAAAGATTCTAAGCTTCGAGTCT
TCGGACTCTTTGGTAACTGAGAGGCCTAAGCATCAATCAAACGAAGTGAAGATAAACGCT
GAACCTGTTTCAACAAGAACCAACAACCTTCAAGGGTTACTTCTCCCCAAGCATCCCCT
GTTTCGGCTCCTTGGCCTTACCAATACCCTCCAAACCTAGTTTCTACCACATGCCCGTC
TACTGGGGCTGCGCGATACCGTTTGGTCTACCCTCGACACTTCTACATGTCTTGGGAAA
AGGACAAGAGACGAAACTTCTCATGAACTGTTAAAGAGAGTAAAAATGCTTTTGAGAGA
ACAAGCTTGCTTTTGAATCTCAGAGCATCAAAAATGAAACAAGTATGGCTACAAATAAC
CATGTGTGGTATCCAGTACCGATGACCCGCGAGAAGACACAAGAATTCAGCTTTTTCAGT
AATGGAGCTGAAACAAAGAGCAGCAACAACAGATTCTGTCCTGAAACGTATCTTAACCTG
CAAGCAAACCTGCAGCCATGGCAAGATCTATGAACCTCAGAGAGAGCATATAA

>G1895 Amino Acid Sequence (domain in AA coordinates: 55-110)

MNNQSVTDNTSLKLSSNLNNEKSEKSENSDDQHSEITITSEEKTELKPKDKILPCPR
CNSADTKFCYNNYNNVNPQRHFCRKCQRYWTAGGSMRIVPVGSGRRKNKGWSSDQYLHI
TSEDTDNYSSTKILSFESSDSLVTTERPKHQSNEVKINAEPVSQEPNNFQGLLPQASP
VSPWPYPYPPNPSFYHMPVYWGCAIPVWSTLDTSTCLGKRTRDETSKETVKESKNAFER
TSLLESQSIKNETSMATNNHVWYPVPMTREKTQEFSSFGAETKSSNNRFPETYLNL
QANPAAMARSMNFRESI*

>G1900 (1..897)

ATGCTGGAAACTAAAGATCCTGCGATAAAGCTCTTTGGTATGAAAATTCCTTTCCCGACG
GTTTTAGAGGTTGCTGATGAAGAAGAAGAAAAGAACCAACAAAGACATTAAGTATCAA
TCGGAGAAAGACAAACCTTAAAGAAACCAACCAAGATTCTTCCATGTCCAAGATGCAAC
AGCATGGAGACTAAGTTCTGTACTACAACAACCTACAACGTAACCAACCTCGCCATTTT
TGTAAGCTTGTGAGAGATATTGGACCTCAGGTGGGACCATGAGAAGTGTCCAATCGGA
GCAGGACGGCGCAAGAACAAGAACAACTCACCACCTTACATTACCACCATGTGACTATC
TCCGAAACAAATGGTCCGGTCTTAGTTTCAGCCTCGGAGATGATCAAAGGTCTCGAGT
AATAGTTTGGTAAATCAAAGCTAGTTGCTAGGATAGAGAACAATGACGAGCGCTCTAAT
AACAACACTTCGAACGGTTTGAATGTTTCCGGGAGTTTCGTGGCCGTACACGTGGAAT
CCTGCGTTTTACCCGTTTTACCCCTTATTGGAGCATGCCAGTGTTGTCTTCTCCGGTAAGT
TCAAGTCTACTTCTACTCTTGGTAAGCATTGAGAGACGAAGACGAGACGGTGAAGCAA
AAACAGAGGAATGGATCTGTATTGGTTCCAAAGACTTTGAGAATTGATGATCCTAATGAA

GCTGCAAAGAGTTTCGATATGGACAACACTTGGGATCAAGAACGAAGTTATGTTCAATGGG
 TTTGGTTCGAAGAAAGAGGTTAAGCTCAGTAACAAAGAAGAAACAGAGACCTCACTTGTT
 CTTTGTGCAAACCTGCTGCGTTATCAAGATCAATCAATTCCATGAGCAGATGTGA
 >G1900 Amino Acid Sequence (domain in AA coordinates: 54-106)
 MLETKDPAIKLFGMKIPFPTVLEVADEEEKNQNKLTLDQSEKDKTLKKPTKILPCPRCN
 SMETKFCYNNYNVNPQPRHFCKACQRYWTSGGTMRSPVIGAGRRKNKNSPTSHTYHHVTI
 SETNGPVLSPSLGDDQKVSSNRFNGNQLVARIENNDERSNNNTSGLNCFPGVSWPYTWN
 PAFYPVPYWSMPVLSSPVSSSPTSTLGKHSRDEDETVKQKQRNGSVLVPKTLRIDDPNE
 AAKSSIWTTTLGIKNEVMFNGFGSKKEVKLSNKEETETSLVLCANPAALSRSINFHEQM*
 >G2007 (1..861)
 ATGGGAAGGCAGCCATGTTGTGACAAGCTCATGGTGAAGAAGGGCCGTGGACGGCGGAG
 GAAGACAAGAACTGATAAACTTTATCTTGACCAACGGCCACTGTTGCTGGAGGGCTTGT
 CCGAAGCTGGCCGCTCTCCGTCGCTGTGGGAAGAGCTGCCGTCTACGGTGGACCAATTAT
 CTCCGACCTGACTTGAAGAGAGGTCTTCTCTCCGACGCCGAGGAACAGCTTGTTCATCGAC
 CTTTCATGCTCTTCTCGGCAACAGATGGTCCAAGATCGCTGCAAGATTACCAGGAAGAACA
 GACAACGAAATAAAAAATCATTGGAATACTCATATCAAGAAGAAGCTCCTTAAGATGGAA
 ATCGATCCTTCGACCCATCAACCTTTAAACAAAGTATTTACCGATACAAACTTAGTCGAT
 AAATCTGAACTTCATCGAAAGCCGACAATGTAAATGATAATAAAATCGTAGAGATCGAT
 GGGACAACGACAAATACAATAGATGATAGCATTATCACTCATCAAAATAGTTCAAATGAT
 GATTATGAATTACTTGGTGATATAATTTCATAATTATGGAGATTATTTAATATTCTATGG
 ACCAACGATGAACCTCTCTAGTCGATGATGCATCATGGAGCAATCATAACGTTGGTATT
 GGAGGAACAGCTGCAGTTGCAGCCTCAGACAAGAACAACACTGCTGCCGAGGAAGATTTTC
 CCGGAAAGATCATTTGAAAAACAGAACGGCGAAAGTTGGATGTTCTTGGATTATTGCCAA
 GAATTTGGTGTTGAAGATTTTGGGTTTCGAGTGTTACCATGGTTTTGGTCAAAGCTCCATG
 AAGACGGGTCACAAGGACTAG
 >G2007 Amino Acid Sequence (domain in AA coordinates: TBD)
 MGRQPCCDKLMVKGPWTAEEDKKLINFILTNHGCCWRALPKLAGLRRCGKSCRLRWNTY
 LRPDLKRGLLSDAEEQLVIDLHALLGNRWSKIAARLPGRDNEIKNHWNTHIKKLLKME
 IDPSTHQPLNKVFTDTNLVDKSETSSKADNVNDNKIVEIDGTTTNTIDDSIITHQNSSND
 DYELLDGI IHNHYGDLFNILWTNDEPPLVDDASWSNHNVGIGGTAAVAASDKNNTAAEEDF
 PERSFEKQNGESWMFLDYCQEFGEVDFGFECYHGFQSSMKTGHKD*
 >G214 (238..2064)
 TGAGATTTCTCCATTCCGTAGCTTCTGGTCTCTTTTCTTTGTTTCATTGATCAAAAGCA
 AATCACTTCTTCTTCTTCTTCTTCTCGATTCTTCTACTGTTTTCTTATCCAACGAAATCTG
 GAATTAAAAATGGAATCTTTATCGAATCCAAGCTGATTTTGTTCCTTTTCATTGAATCATC
 TCTCTAAAGTGGAATTTTGTAAAGAGAAGATCTGAAGTTGTGTAGAGGAGCTTAGTGATG
 GAGACAAATTTCGTCTGGAGAAGATCTGGTTATTAAGACTCGGAAGCCATATACGATAACA
 AAGCAACGTGAAAGGTGGACTGAGGAAGAACAATAAGATTTCATTGAAGCTTTGAGGCTT
 TATGGTAGAGCATGGCAGAAAGATTGAAGAACATGTAGCAACAAAACTGCTGTCCAGATA
 AGAAGTCACGCTCAGAAATTTTCTCCAAGGTAGAGAAAGAGGCTGAAGCTAAAGGTGTA
 GCTATGGGTCAAGCGCTAGACATAGCTATTCTCTCTCCACGGCCTAAGCGTAAACCAAAC
 AATCCTTATCTCGAAAGACGGGAAGTGAACGATCCTTATGTCAAAAACGGGTGTGAAT
 GATGGAAAAGAGTCCCTTGGATCAGAAAAAGTGTGCGATCCTGAGATGGCCAATGAAGAT
 CGACAACAATCAAAGCCTGAAGAGAAAACTCTGCAGGAAGACAACCTGTTTCAGATTGTTTC
 ACTCATCAGTATCTCTCTGCTGCATCCTCCATGAATAAAAAGTTGTATAGAGACATCAAAC
 GCAAGCACTTTCCGCGAGTTCTTGCCTTCACGGGAAGAGGGAAGTCAGAATAACAGGGTA
 AGAAAGGAGTCAAACCTCAGATTGAATGCAAAATCTCTGGAAAACGGTAATGAGCAAGGA
 CCTCAGACTTATCCGATGCATATCCCTGTGCTAGTGCCATTGGGGAGCTCAATAACAAGT
 TCTCTATCACATCCTCCTTCAGAGCCAGATAGTCATCCCCACACAGTTGCAGGAGATTAT
 CAGTCGTTTCTTAATCATATAATGTCAACCCTTTTACAAACACCGGCTCTTTATACTGCC
 GCAACTTTTCGCTCATCATTTTGGCCTCCCGATTCTAGTGGTGGCTCACCTGTTCCAGGG
 AACTCACCTCCGAATCTGGCTGCCATGGCCGAGCCACTGTTGCGAGCTGCTAGTGCTTGG
 TGGGCTGCCAATGGATTATTACCTTTATGTGCTCCTCTTAGTTTCAGGTGGTTTCACTAGT
 CATCCTCCATCTACTTTTGGACCATCATGTGATGTAGAGTACACAAAAGCAAGCACTTTA
 CAACATGGTTCTGTGCAGAGCCGAGAGCAAGAACAACCTCCGAGGCATCAAAGGCTCGATCT
 TCACTGGACTCAGAGGATGTTGAAAATAAGAGTAAACCAGTTTGTTCATGAGCAGCCTTCT
 GCAACACCTGAGAGTGATGCAAAGGGTTCAGATGGAGCAGGAGACAGAAAACAAGTTGAC

CGGTCCTCGTGTGGCTCAAACACTCCGTCGAGTAGTGATGATGTTGAGGCGGATGCATCA
GAAAGGCAAGAGGATGGCACCAATGGTGAGGTGAAAGAAACGAATGAAGACACTAATAAA
CCTCAAACCTCAGAGTCCAATGCACGCCGAGTAGAATCAGCTCCAATATAACCGATCCA
TGGAAGTCTGTGTCTGACGAGGGTGAATTGCCTTCCAAGCTCTCTTCTCCAGAGAGGTA
TTGCCGCAAAGTTTTACATATCGAGAAGAACACAGAGAGGAAGAACAACAACAAGAA
CAAAGATATCCAATGGCACTTGATCTTAACCTCACAGCTCAGTTAACACCAGTTGATGAT
CAAGAGGAGAAGAGAAACACAGGATTTCTTGGAATCGGATTAGATGCTTCAAAGCTAATG
AGTAGAGGAAGAACAGGTTTTAAACCATACAAAAGATGTTCCATGGAAGCCAAAGAAAGT
AGAATCCTCAACAACAATCCTATCATTCATGTGGAACAGAAAGATCCCAAACGGATGCGG
TTGGAAACTCAAGCTTCCACATGAGACTCTATTTTCATCTGATCTGTTGTTGTAATCTG
TTTTTAAGTTTTCAAGACCCTGCTACATTTCTTTTCTTTTGAGGCCTTGTATTTGT
TTCCTTGTCATAGTCTTCTGTAAACATTTGACTCTGTATTATTCAACAAATCATAAACT
GTTTAATCTTTTTTTTCCCA

>G214 Amino Acid Sequence (domain in AA coordinates: 22-71)
METNSSGEDLVIKTRKPYTITKQREERWTEEEHNRFIELRLYGRAWQKIBEHVATKTAVQ
IRSHAQKFFSKVEKEAEAKGVAMGQALDIAIPPPRPKRKPNNPYPRKTGSGTILMSKTGV
NDGKESLGSEKVSHPMANEDRQQSKPEEKLQEDNCSDFTHQYLSAASSMNKSCIETS
NASTFREFLPSREEGSQNNRVRKESNSDLNAKSLENGNEQGPQTYPMHIPVLVPLGSSIT
SSLSHPPSEPDSHPHTVAGDYQSFPHIMSTLLQTPALYTAATFASSFWPPDSSGGSPVP
GNSPPNLAAMAAATVAASAWWAANGLLPLCAPLSSGGFTSHPPSTFGPSCDVEYTKAST
LQHGVSQSREQEHSEASKARSSLDSEDEVENKSKPVCHEQPSATPESDAKGSAGDRKQV
DRSSCGSNTPSSDDVEADASERQEDGTNGEVKETNEDTNKPQTSSESNARRSRISNITD
PWKSVSDEGRIFAFQALFSREVLQSFYTYREEHREEEQQQEQRYPMALDLNFTAQLTPVD
DQEEKRNTGFLGIGLDASKLMSRGRGTGFKPKYKRCSMEAKESRILNNNPIIHVEQKDPKRM
RLETQAST*

>G2155 (63..740)

CTCATATATACCAACCAAACCTCTCTGTCATCTTTATTAAACACAAAATTCCAAAAGATT
AAATGTTGTGCGAAGCTCCCTACACAGCGACACTTGACCTCTCTCCCTCCTCTCCCTCCA
TGGAACCGTCCGGGCGTCCACGTGGCAGACCTCGAGGTTCCAAAACAAACCTAAAGCTC
CAATCTTTGTGTCACCATTTGACCTCCTATGAGTCCCTACATCCTCGAAGTGCCATCCGGAA
ACGATGTGCTGTTGAAGCCCTAAACCGTTTCTGCCGCGGTAAAGCCATCGGCTTTTGCGTCC
TCAGTGGCTCAGGCTCCGTTGCTGATGTCACTTTGCGTCAGCCTTCTCCGGCAGCTCCTG
GCTCAACCATTACTTTCCACGGAAAGTTGATCTTCTCTGTCTCCGCCACTTTTCTCTCC
CTCTCTACCTCCTACCTCCTTGTCCCCTCCCGTCTCCAATTTCTTCACCGTCTCTCTCG
CCGGACCTCAGGGGAAAGTCATCGGTGGATTCTGTCGCTGGTCTCTCGTTGCCGCCGGAA
CTGTTTACTTCGTCGCCACTAGTTTCAAGAACCCTTCTATACCGGTTACCTGCTACGG
AGGAAGAGCAAAGAACTCGGCGGAAGGGGAAGAGAGGGACAATCGCCGCCGGTCTCTG
GAGGTGGTGGAGAGTCGATGTACGTGGGTGGCTCTGATGTCAATTTGGGATCCCAACGCCA
AAGTCCATCGCCGTACTGACCACAAATCCATCTCGTTCAAACAGGGTTTCTTCTTCTT
TAGATCATCAAGAATCAACAAAAGATTGCATTTTAGATTCTTTGTAATATCATAATTG
ACTCACTCTTTAATCTCTCTATCACTTCTTCTTTAGCTTTTCTGTCAGTGTCAAACCTCA
CATATTTGTAGTTTGTATTTGACTATCCCCAAGTTTGTATTTTATCATACAAATTTTGC
CTGTCTCTAATGGTTGTTTTTTCGTTTGTATAATCTTATGCATTGTTTATTGGAGCTCCA
GAGATTGAATGTATAATATAATGGTTTAAT

>G2155 Amino Acid Sequence (domain in AA coordinates: 18-38)
MLSKLPTQRHLHLSPPSPSMETVGRPRGRPRGSKNPKAPIFVTIDPPMSPYILEVPSGN
DVVEALNRFRCRKAIGFCVLSGSGSVADVTLRQPSAPAGSTITFHGKFDLLSVSATFLP
PLPPTSLSPPVSNFPTVSLAGPQGVIGGFVAGPLVAAGTVYFVATSFKNPSYHRLPATE
EEQRNSAEGE EEGQSPPVSGGGESMYVGGSDVIWDPNAKAPSPY*

>G234 (106..1035)

CACAACATCATACCCACCAACATATATAATCTTGATCATAGAGAGATAAAACAGAGGCCGC
TATCAAGAACAAGACTAAGAACAAGACTTCACTAGGAGTACAAGTATGGGAAGAGCACCG
TGTTGTGACAAAGCAAACCTGAAGAAAGGGCCTTGGTCTCCTGAGGAAGATGCAAACTC
AAATCTTACATTGAAAATAGTGGCACCGGAGGCAATTGGATCGCTTTCCTCAAAGATT
GGTTTAAAGAGATGTGGAAAGAGTTGCAGGCTGAGGTGGCTTAACTATCTTAGACCAAAC
ATCAAACATGGTGGCTTCTCTGAGGAAGAAGAAAACATCATTTGTAGCCTTTACCTTACA
ATTGGTAGCAGGTGGTCTATAATCGCTGCTCAATTGCCGGGACGAACAGACAACGATATA

AAAACTATTGGAACACGAGGCTCAAGAAGAACTCATTAACAAACAACGCAAGGAGCTT
CAAGAAGCTTGTATGGAGCAGCAAGAGATGATGGTGATGATGAAGAGACAACACCAACAA
CAACAAATCCAACTTCTTTTATGATGAGACAAGACCAACAATGTTTACATGGCCACTA
CATCATCATAATGTTCAAGTTCAGCTCTTTTCAGAATCAAACCAACTCGTTTTGCGACC
AAGAAGATGTTAAGCCAGTGCTCATCAAGAACATGGTCAAGATCGAAGATCAAGAAGTGG
AGAAAACAAACCTCATCATCATCAAGATTCAATGACAACGCTTTTGATCATCTCTCTTTC
TCTCAACTCTTGTAGATCCTAATCATAACCACTTAGGATCAGGAGAGGGTTTCTCCATG
AACTCTATCTTGAGCGCCAACACAACTCTCCATTGCTTAACACAAGTAATGATAATCAG
TGGTTCCGGAATTTCCAGGCCGAAACCGTAAACTTGTCTCAGGAGCCTCCACAAGTACT
TCGGCAGATCAAAGCACTATAAGTTGGGAAGACATAAGCTCTCTTGTATTCTGATTCA
AAGCAATTTTTTTAATTATAATAATATATTCTTAAGATGAAACGTACATCATTATTA
TTAATTGGGGGTACGTAACGTATATATGGAATAACGATCTAGTTTGTTTAAATTTAAAA
>G234 Amino Acid Sequence (domain in AA coordinates: 14-115)
MGRAPCCDKANVKKGPWSPEEDAKLSYIENSGTGGNWIALPQKIGLKRCKGKSCRLRWLN
YLRPNIKHGGFSEEEENIICSLYLTIGSRWSIIAQLPGRDNDIKNYWNTRLKKLINK
QRKELQEACMEQQEMVMKROHQQQQIQTSFMMRQDQTMFTWPLHHHNQVQVPALFRIKP
TRFATKMLSQCSSRTWSRSKIKNWRKQTSSSSRFNDNAFDHLSFSQLLDPNHNHLGSG
EGFSMNSILSANTNSPLNLSNDNQWFGNFAETVNLFSGASTSTSDQSTISWEDISSL
VYSDSKQFF*

>G361 (54..647)

TCTGTCTCTCTCTCTCTCTTTGTAAATATACATATATAGATAAGCTCACATATATGGCGA
CTGAAACATCTTCTTTGAAGCTCTTCGGTATAAACCTACTTGAAACGACGTCGGTTCAAA
ACCAGTCATCGGAACCAAGACCCGGATCCGGATCAGGATCCGAGTCACGTAAGTACGAGT
GTCAATACTGTTGTAGAGAGTTTGCTAACTCTCAAGCTCTTGGTGGTCACCAAAACGCTC
ACAAGAAAGAGCGTCAGCTTCTTAAACGTGCACAGATGTTAGCTACTCGTGGTTTGCCAC
GTCATCATAATTTTCACCCCTACATCAATCCGCTTCTCTCCGCTTCGCGCCGCTGCCTC
ACCTCTCTCTCAGCCGCATCCTCCGCCGATATGATGCTCTCTCTCTCTCTCTCTCGAGTT
CTAAGTGGCTTTTACGGAGGTGACACATGTCGTACAAAACGCCGTTGGGTACTTTCATGGTG
GAAGGGAGCTTTACGGAGGTGGCATGGAGTCTATGGCCGGAGAAGTAAAGACTCATGGTG
GTTCTTTGCCGAGATGAGGAGGTTCCGCCGAGATAGTGATCGGAGTAGCGGAATTAAGT
TAGAGAATGGTATTGGGCTGGACCTCCATTTAAGCCTTGGGCCATGAATGATTATAATTT
TGGCCAGTAAAGATCTGTAAATACTACTAGGATTTTATTTTATAGAGTATGTTTTTT
TCCTTAATTTTCGGTTGAAATTGGTGAATATTTTATCTCTTACTTACCAATCTCATATT
TCTATGTATGCGTTTGCTTCACTTTTTTTTTTATATAATTCTTCTGTAAAAAATGCA
ATGTGAGTTTTCTTCCCTATCATCTGTCAAGCTTTGGTTCAATTATTTAGTAATCGAAT
AATATAGGAATAGTGTGAAAG

>G361 Amino Acid Sequence (domain in AA coordinates: 43-63)

MATETSSLKLFGINLLETTSVQNQSSEPRPGSGSSESRYECQYCCREFANSQALGGHQ
NAHKKERQLLKRAQMLATRGLPRHNFHPHTNPLLSAFAPLPHLLSQPHPPPHMMLSPSS
SSSKWLYGEHMSSQNAVGYFHGGRGLYGGMESMAGEVKTHGGSLPEMRRFAGDSDRSSG
IKLENGIGLDLHLSLGP*

>G562 (137..1285)

ATTTGAATTTCTGGGTTTCTCTCTGTTTAAGCTTCTTCTTCTTCTTCTGCTTACGTT
TCTTCTTCAAGGAGCTTTCGGATTCTTGTAGAAAGAGTCATTGTTCTCTTGAGTGGGAAA
CCTTGAAACCATCTTCTATGGGAAATAGCAGCGAGGAACCAAGCCTCCTACCAAATCAGA
TAAACCATCTTTCACCCCGGTGGATCAAACAAATGTTATGTCTACCTGATTGGGCGAG
TATGCAGGCATATTATGGTCCAAGAGTAGCAATGCCTCCTTATTACAATTAGCTATGGC
TGCATCTGGTCATCTCTCTCTCTTACATGTGGAATCCTCAGCATATGATGTCACCATC
TGGAGCACCTATGCTGCTGTTTATCCTCATGGAGGAGGAGTTTACGCTCATCCCGGTAT
TCCCATGGGATCACTGCCTCAAGGTCAAAGGATCCACCTTTAACAACCTCCGGGGACGCT
TTTGAGCATCGACACTCTTAAATCTACAGGGAACACAGACAATGGATTGATGAAGAA
GCTGAAAGAGTTTGATGGGCTTGCTATGTCTCTAGGAAATGGGAATCCTGAAAATGGTGC
AGATGAACATAAACGATCACGGAACAGCTCAGAACTGATGGTTCTACTGATGGAAGTGA
TGGGAATACAACTGGGGCAGATGAACCGAACTTAAAGAAGTCGAGAGGGAACCTCCAAC
AAAAGATGGGAAACAATTGGTTCAAGCTAGCTCATTTTCTGTTTCTCCGTCAAGTGG
TGATACCGGCGTAAACTCATTCAGGATCTGGAGCTATACTCTCTCTGTTGTAAGTGC
AAATTCACACCCCTTCATGTCAATCTTTAGCCATGGTTCCTCCTGAAACTTGGCTTCA

GAACGAGAGAGAACTGAAACGGGAGCGAAGGAAACAGTCTAATAGAGAATCTGCTAGAAG
GTCAAGATTAAGGAAACAGGCCGAGACAGAAGAACTTGCTAGGAAAGTGGAAGCCTTGAC
AGCCGAAAACATGGCATTAAAGATCTGAACTAAACCAACTTAATGAGAAATCTGATAAACT
AAGAGGAGCAAATGCAACCTTGTGTGGACAACTGAAATGCTCGGAACCCGAAAAGAGAGT
CCCCGCAAATATGTTGTCTAGAGTTAAGAACTCAGGAGCTGGAGATAAGAACAAGAACCA
AGGAGACAAATGATTTCTAACTCTACAAGCAAATTCCATCAACTGCTCGATACGAAGCCTCG
AGCTAAAGCAGTAGCTGCAGGCTGAATCGATGGTAATTCATGTCTGATTCTACTTAATTT
GTCGACATAAAACAAAGAAAATAAGTGCTACTAATTTTCAGAAAACTTGATAGATAGATAG
TATAGTAGAGAGAGAGAGAGAGAGAGAGGTGTGATGATTATTGATCTATAAATTTTCGGA
GAGAGAGAGGGAGAAAGAGAACTTTTCCTCCAGATGAAAATTTGGTGTATGGTTGTT
ACTGTTAATATAGAGAGGCTTTTCTTTTATAAAAATGGCTTCCTTTGTTGCA
>G562 Amino Acid Sequence (domain in AA coordinates: 253-315)
MGNSSEEPKPTKSDKPSPPVDQTNVHVPDWAAMQAYYGPVAMPPIYNSAMAAAGHP
PPPYMWNQPHMMSPSGAPYAAYVPHGGGVYAHPGIPMGSPLPQGGKDPPLTTPGTLLSIDT
PTKSTGNNDNGLMKKLKEFDGLAMSLGNGNPENGADCHKRSRNSSETDGDSDGSDGNTTG
ADEPKLKRREGTPTKDGKQLVQASSFHSVSPSSGDTGVKLIQSGAILSPGVSANSNPF
MSQSLAMVPPETWLQNERELKRERRKQSNRESARRSRLRQAETEELARKVEALTAENMA
LRSELNQLNEKSKDLRGANATLLDKLKCSEPEKRV PANMLSRVKN SGAGDKNKNQGDNDNS
NSTSKFHQLLDTKPRAKAVAAG*
>G591 (88..1020)
GTAAATCTCTCTTTGAAGGTTCTTAACCTCGTTAATCGTAACTCACAGTGACTCGTTTCGAG
TCAAAGTCTCTGTCTTTAGCTCAAACCATGGCTAGTAACAACCCCTCACGACAACCTTTCT
GACCAAACCTCTTCTGATGATTCTTCGAGCAAATCCTCGGCCTTCTTAACCTCTCAGCC
TCTTCTGCCGCCGGTTTATCTGGAGTTGACGGAGGATTAGGTGGTGGAGCACC GCCTATG
ATGCTGCAGTTGGGTTCCGGAGAAGAAGGAGTCACATGGGTGGCTTAGGAGGAAGTGGA
CCAACCTGGGTTTCACAATCAGATGTTTCTTTGGGGTTAAGTCTTGATCAAGGGAAGGA
CCTGGGTTTCTTAGACCTGAAGGAGGACATGGAAGTGGGAAAAGATTCTCAGATGATGTT
GTTGATAATCGATGTTCTTCTATGAAACCTGTTTTCCACGGGCAGCCTATGCAACAGCCA
CCTCCATCGGCCCCACATCAGCCTACTTCAATCCGTCACGGGTTTCGAGCTAGGCGTGGT
CAGGCTACTGATCCACATAGCATCGCTGAGCGGTACGTAGAGAAAGAATAGCAGAACGG
ATCAGGGCGCTGCAGGAACCTTGTAACCTACTGTGAACAAGACCGATAGAGCTGCTATGATC
GATGAGATTGTGCTGATTATGTAAAGTTTCTCAGGCTCCAAGTCAAGGTTTTGAGCATGAAC
CGACTTGGTGGAGCCGGTGCGGTTGCTCCACTTGTACTGATATGCCTCTTTCATCATCA
GTTGAGGATGAAACCGGTGAGGTGGAAGGACTCCGCAACCAGCGTGGGAGAAATGGTCT
AACGATGGGACTGAACGTCAAGTGGCTAAACTGATGGAAGAGAACGTTGGAGCCCGATG
CAGCTTCTTCAATCAAAGGCTCTTGTATGATGCCAATCTCATTGGCAATGGCAATTTAC
CATTCTCAACCTCCGGATACATCTTCAAGTGGTCAAGCCTGAGAACAAATCTCCACAGTAG
GATTTCTGCAATAAAGAGTTTGTACAGCTAATCCAACGTCCAACATGGGTTTTTCTTCT
GCTCTAATGACTCTGGTTTTCTTCTCTCTCTCACCAGCTTGAAAGGTAAAAAAGTGAA
AAAGGCTTTGTAGATGGAATCAATGTAGGATTTGCAGTAGAGGGCAAAAAATGTCATAT
AGCTCAATTGATCAAGTCTTAAAAA AAAAAAAAAAAAAAAAAA
>G591 Amino Acid Sequence (domain in AA coordinates: 143-240)
MASNNPHDNLSDQTPSDDFFEQILGLPNFSASSAAGLSGVDGGLGGGAPPMMLQLGSCEE
GSHMGGGLGSGPTGFHNQMFPLGLSLDQKGPGFLRPEGGHGSGKRFSDDVVDNRCSSMK
PVFHGQPMQPPPSAPHQPTSIRPRVRARRGQATDPHSIAERLRRERIAERIRALQELVP
TVNKTDRAMIDEIVDYVKFLRLQVKVLSMNLGAGAVAPLVTDMPLSSSVEDETGE
RTPQPAWEKWSNDGTERQVAKLMEENVGAAMQLLQSKALCMMPISLAMAIYHSQPPDTSS
VVKPENPPQ*
>G8 (247..1596)
AAAAAAAAAATATCCGTCTCACTCTCTCGCCGCCGTAACATTTCCCGGCACAAAACCTTC
TCTACTCTCACCATTCTCCATCGTAATCTCTAAATCTTCTCCATTCTTCTTCTCTCC
CGATCATCTCGAGCTCTTCGTGAGAGATTATGTGATTATGTAATCGTTGTGCTGTAGAA
GACGATCTCTAACAACCTGATTCCTTCATCATCACCTTCGCTAGATTTGTAATTTTCAGAG
CTTGAGATGTTGGATCTTAACCTCAACGCTGATTCTCCCGAGTCGACTCAGTACGGTGGT
GACTCATACTTAGATCGGCAGACATCAGACAACTCCGCCGGGAATCGAGTGAAGAGTCC
GGTACATCGACGTCGTCAAGTTATCAATGCCGATGGAGACGAAGACTCTTGCTCTACTCGA
GCTTTCACTCTCAGTTTCGATATTTTAAAGTCGGAAGTAGTAGCGGCGGAGACGAAAGC

CCCGCCGCTTCAGCTTCCGTTACTAAAGAGTTTTTCCGGTGAGTGGAGACTGTGGACAT
 CTACGAGATGTTGAAGGATCATCAAGCTCTAGAACTGGATAGATCTTCTTTTGACCGT
 ATTGGTGACGGAGAAACGAAATTGGTAACTCCGGTCCGACTCCGGCTCCGGTCCGGCT
 CAGGTTAAAAAGAGTCGGAGAGGACCAAGGTCTAGAAGTTCACAGTATAGAGGAGTTACT
 TTTTATAGAAGAACTGGTCGATGGGAGTCACATATTTGGGATGTGGGAAACAAGTTTAT
 TTAGGTGGTTTCGACACTGCTCATGCTGCAGCTAGAGCTTATGATCGAGCTGCTATTAAA
 TTTAGAGGTGTTGATGCTGATATCAACTTTACTCTTGGTGATTATGAGGAAGATATGAAA
 CAGGTACAAAACTTGAGTAAGGAAGAGTTTGTGCATATACTGCGTAGACAGAGCACGGGG
 TTTTCGCGGGGAGTTCGAAGTATCGAGGGGTTACGTTACACAAATGTGGTAGATGGGAA
 GCTAGGATGGGGCAGTTTCTTGGTAAAAAGGCTTATGACAAGGCTGCAATCAACACTAAT
 GGTAGAGAAGCAGTCACGAACCTCGAGATGAGTTCATACCAAAATGAGATTAACTCTGAG
 AGCAATAACTCTGAGATTGACCTCAACTTGGGAATCTCTTTATCGACCGGTAATGCGCCA
 AAGCAAAATGGGAGGCTCTTTCACTTCCCTTCTAATACTTATGAAACTCAGCGTGGAGTT
 AGCTTGAGGATAGATAACGAATACATGGGAAAGCCGGTGAATACACCTCTTCTTATGGA
 TCCTCGGATCATCGCCTTTACTTGGAACGGAGCATGCCCGAGTTATAATAATCCCGCCGAG
 GGAAGAGCAACAGAAAAGAGAAGTGAAGCTGAAGGGATGATGAGTAACGGGGATGGCAG
 AGACCGGGGCAACAAGCGCCGTGAGACCGCAGCCACCGGGACCACAACCACCACCATG
 TTCTCAGTTGCAGCAGCATCATCAGGATTCTCACATTTCCGGCCACAACCTCCCAATGAC
 AATGCAACACGTGGTTACTTTTATCCACACCCTTAACTTGTAAGGGGACATATGAGAGTT
 TTTTACCATCTCTCTCTCTCAACACTCTAGTCCCTTTCAAAAATGTCATTGGGGT
 TTAGATTTTTCACATACAATGATCAATTTTTC

>G8 Amino Acid Sequence (domain in AA coordinates: 151-217, 243-296)

MLDLNLNADSPSTQYGGDSYLDROTSDNSAGNRVEESGTSTSSVINADGDEDSCTRAF
 TLSFDILKVGSSSGGDESPAASASVTKEFFPVSGDCGHLRDVEGSSSSRNWIDLSFDRIG
 DGETKLVTVPPTAPVPAPQVKSSRRGPRSRSSQYRGVTFYRRTGRWESHIWDCGKQVYLG
 GFDTAHAARAYDRAAIKFRGVDADINFTLGDYEEDMKQVQNLKKEEFVHILRRQSTGFS
 RGSSKYRGVTLHKCGRWEARMGQFLGKKAYDKAAINTNGREAVTNFEMSSYQNEINSESN
 NSEIDLNLGISLSTGNAPKQNGRLFHFPSTYETQRGVSLRIDNEYMGKPVNTPLPYGSS
 DHRLYWNGACPSYNNPAEGRATEKRSEAEGMMSNWGWQRPQTSAPVRPQPPGPQPPPLFS
 VAAASSGFSHFRPQPPNDNATRGYFYPHP*

>G859 (162..752)

GATTTGTCAATTTTTGTCTAGCCAAAAAAAAAAAAAAAAAGGAGAGAGAGAGAGAGAGA
 GAGAGAGAGAGAAACGAAGAAAAAAAAAGAGCAAAAAACATTGTGGGTCTCCGGTGATT
 AGGATCAAAATTAGGGCACCAGCCTTATCGGAGGAAGAAGCCATGGGTAGAAAAAAGTCG
 AGATCAAGCGAATCGAGAACAAAAGTAGTCGACAAGTCACTTTCTCCAAACGACGCAATG
 GTCTCATCGAGAAAGCTCGACAACCTTCAATTCTCTGTGAATCTTCCATCGCTGTCTCG
 TCGTCTCCGGCTCCGGAAAACTCTACAAGTCTGCCTCCGGTGACAACATGTCAAAGATCA
 TTGATCGTTACGAAATACATCATGCTGATGAACCTGAAGCCTTAGATCTTGCAAAAAA
 CTCGGAATTATCTGCCACTCAAAGAGTTACTAGAAATAGTCCAAAGCAAGCTGAAGAAT
 CAAATGTCGATAATGCAAGTGTGATACTTTAATTTCTCTGGAGGAACAGCTCGAGACTG
 CTCTGTCCGTAAGTAGAGCTAGGAAGACAGAACTAATGATGGGGGAAGTGAAGTCCCTTC
 AAAAAACGGAGAACTTGCTGAGAGAAGAGAACCAGACTTTGGCTAGCCAGGTGGGGAAGA
 AGACGTTTCTGGTTATAGAAGGTGACAGAGGAATGTCATGGGAAAATGGCTCCGGCAACA
 AAGTACGGGAGACTCTTCCGCTGCTCAAGTAATCACCATCATCAACGGCTGAGCTTTCAC
 CTTAAACTTACAGCCTGATTCAAGAGTTTTTACAAATTTGTAAATTATAAAAAAGCTTCAT
 AATAATCTCAACCTTTTTATCTTCTCGCGCAATGTGGAAATTAAGGTTAAAAATAAAA
 TAAACAGAAGCTCATGCGAAAGAATTGTAAACTAAGATAAAGCTATAGTAGATCTTTA
 TTGTACCTTCGTAGACGATATAAGATTTATTCGTGTGTTGTCTTCCCTCNAAAAAAAA
 AAAAAAAAAAAAAA

>G859 Amino Acid Sequence (domain in AA coordinates: TBD)

MGRKKVEIKRIENKSSRQVTFSKRRNGLIEKARQLSILCESSIAVLVSGSGKLYKSASG
 DNMSKIIDRYEIIHHADELEALDAEKTRNYLPLKELLEIVQSKLEESNVDNASVDTLISL
 EEQLETALSVTRARKTELMMEGVKSLQKTENLLREENQTLASQVGKKTFLVIEGDRGMSW
 ENGSGNKVRETLPPLK*

>G878 (197..1738)

CAAAAAATCTCTCCATTAAAAGACTGCCCAAAGAAATATTTTATACAAAATGAAAGA
 GAGAAACACGACACGAATTTGTATAATTAAGATTACACAAAAAAGTGTTAGAAAGAG

AAATATCTTCTCTTTTTCTGTGTGAGTTGGGTTTGTAAAGTTTATCCTTTTTGTTC
 TCAAAATCAAGAATCGATGGCGGAGAAGGAAGAAAAAGAACCATCGAAGTTAAATCATC
 CACCGGAGTTTACGGCCAACGATTTCACTACCTCCTCGACCGTTTGGTGAAATGTTTTT
 TAGCGGTGGCGTTGGATTTAGTCTCGACCAATGACTCTCGTCTCAAATTTATTCTCTGA
 TCCTGATGAGTTCAAGTCTTTCTCTCAGCTTTTAGCTGGAGCTATGGCTTCTCCGGCGGC
 AGCTGCTGTTGCCGCGCTGCTGTGGTTGCTACTGCTCATCATCAGACACCTGTGAGCTC
 TGTCGGTGATGGCGGTGGAAGCGGTGGTGATGTTGACCCGAGGTTTAAGCAGAGTAGACC
 AACGGGATTGATGATAACTCAACCACCGGGGATGTTTACTGTACCGCCGGGGTTAAGTCC
 GGCTACTCTTTTGGATTCTCCGAGCTTCTTTGGTCTTTTTTACCTCTTCAGGGAACATT
 TGGTATGACACATCAACAAGCTTTAGCACAAGTCACTGCACAAGCAGTTCAAGGCAATAA
 TGTTTCATATGCAGCAATCACAACATCTGAATATCCTTCTTCTACACAACAACAACA
 ACAACAACAAGCTTCATTGACTGAGATTCCATCATTTTCTTCTGCACCTAGGTCTCA
 GATTCGAGCCTCGGTTCAAGAAACATCGCAGGGTCAGAGAGAGACTTCGGAAATATCTGT
 CTTTGAGCATCGGTACAGCCTCAAAATGCTGACAAACCAGCTGATGATGGATACAACCTG
 GCGGAAATATGGGCAGAAGCAAGTGAAGGGGAGCGATTTTCTCGGAGTTATTACAAATG
 TACGCATCCAGCTTGCTCTGTCAAGAAGAAAGTGGAGAGGTCACTCGATGGACAAGTAAC
 GGAAATCATCTACAAGGGTCAACACAATCATGAGCTTCTCAAAAGCGCGGTAACAATAA
 CGGGAGTTGTAAAAGTTCTGATATTGCAAATCAGTTTCAAACAAGTAATAGCAGTCTCAA
 CAAGAGTAAGAGGGACCAGGAAACAAGCCAAGTTACAACAACAGAGCAGATGCTGAAGC
 AAGTGATAGCGAGGAGGTTGGGAATGCAGAGACTAGTGTGGGAGAAAGACATGAGGATGA
 GCCTGATCCCAAGCGAAGAAATACAGAAGTTCCGGTTTCAGAACCAGTTGCTTCATCGCA
 TAGAAGTGTGACAGAGCCTAGGATTATTGTCCAAACGACGAGTGAAGTTGACCTCTTAGA
 TGATGGATATAGGTGGCGCAAGTATGGTCAGAAAGTAGTCAAAGGAAATCCTTATCCGAG
 GAGCTACTATAAGTGTACAACACCAGATTGCGGAGTAAGGAAACATGTAGAGAGAGCAGC
 AACTGACCCAAAAGCTGTTGTAACAACATATGAAGGTAAACATAACCATGATGTTCCAGC
 TGCTAGAACCAGCAGCCATCAGTTAAGACCAAACAATCAACACAACACCTCAACGGTTAA
 CTTCAATCATCAACAGCCTGTTGCACGTTTAAAGGCTTAAAGAAGAGCAAATCACTTGACA
 GAGAAGAAGAATACGACGCGCTTGAGCTTTTGTGAGTTTAAATGAATCTTCTTTTTGGTT
 AATGAACCTGTTTTTGTGCTCAAAACACCACAGGTTTCTCTGGACAGAATCTCTGATA
 TTACAGTTTCAAAGGTATGTTCTTTTATTTTCATGTTGGAATCTTCTGTGAATCTTAAG
 AAGCTTTAGGAGGTATGTAAAAAACAGATTCAAAGTTATGCCCTTATGTGAATCTTTT
 TGTACATGGGATAAACAAAATTTACAGGTATCCTTTTTGTTCTTGTGTAATAAAAAA
 AAAA

>G878 Amino Acid Sequence (domain in AA coordinates:250-305, 415-475)

MAEKEKEPSKLSSTGVSRTISLPPRPFGEMFFSGGVGFSPGPMTLVSNLFSDFDEFK
 SFSQLLAGAMASPAAAVAAAAVVATAHHQTPVSSVGDGGGSGGDVDPKQSRPTGLMI
 TQPPGMFTVPPGLSPATLLDSPSFFGLFSPHQGTGFMTHQQAQVTAQAVQGNVHMQQ
 SQQSEYPSSTQQQQQQQASLTEIPSFSSAPRSQIRASVQETSQGGRETSEISVFEHRS
 QPQNADKPADDGYNWRKYGQKQVKGSDFPRSYKCTHPACPVKKKVERSLDGQVTEIIYK
 GQHNHELPQKRGNNNGSCKSSDIANQFQTSNSSLNKSQRDQETSQVTTTEQMSEASDSEE
 VGNABTSVGERHEDEPDPKRRNTEVRVSEPVASSHRTVTEPRIIVQTTSEVDLLDDGYRW
 RKYGQKVVKGNPYPRSYKCTTPDCGVRKHVERAATDPKAVVTTYEGKHNHDVPAARTSS
 HQLRPNNQHNTSTVNFNHQPPVARLRLKEEQIT*

>G971 (131..1171)

TTTTTTTCTTCCCTCTTTAGAACTCTCTCTCTCTCGTTTTTGACACTTATCCTCTC
 TCTTTTTCTCTCTCCCTCTCTCTCTGGCCGAAAAAGAACAACGTCGTTTATAGCTAA
 AGATTCGATCATGTTGGATCTTAACCTAAAGATCTTTTCTTCTTATAACGAAGATCAAGA
 TCGGAAAGTACCATTAATGATCTCAACCACCGGTGAAGAAGAATCTAACTCATCTTCCTC
 CTCACAACAGACTCTGCAGCGAGAGATGCTTTCATCGCTTTTGGAAATCTCAAACGCGA
 CGATGACCTTGTTCTCTCTCTCTCTCTCATAAAGAAACAGGAGATCTCTTTCC
 GGTGGTGGCTGATGCTCGTCGGAATATAGAATTTCTCCGTGGAAGACAGTCACTGGTTGAA
 TCTTTCTTCTTTACAAAGAAATACACAGAAAATGGTGAAGAAGAGCAGAAGAGGACCAAG
 GTCTCGTAGCTCCCAATATCGTGGCGTCACTTTTTACCGTCGCACCGGTCGTTGGGAATC
 TCATATTTGGGATTGTGGAAGCAAGTTTATTTGGGCGGGTTTGATACTGCTTACGCAGC
 AGCAAGGGCTTACGACCGAGCTGCTATCAAATCCCGTGGTCTCGATGCAGACATCAATTT
 CGTCGTGGATGATTATAGGCATGACATCGATAAGATGAAGAATTTAAATAAGGTGGAGTT
 CGTGCAAAACACTTAGGCGAGAGAGTGCAGTTTCGGAAGAGGAAGTTCCAAATACAAAGG

CTTGGCTCTTCAAAAATGCACCCAATTCAAACTCATGATCAGATTCTCTCTTCCAAAA
 CAGGGGATGGGATGCAGCAGCAATAAAATACAATGAGTTGGGAAAGGGAGAAGGAGCCAT
 GAAGTTTGGTGCCCATATCAAAGGAAATGGTCACAATGATCTTGAACCTAAGTCTCGGAAT
 TTCATCATCATCGGAAAGTATAAAGTTGACAACAGGCGATTACTATAAGGGTATCAATCG
 GTCCACGATGGGTTTATACGGTAAGCAATCATCGATATTTTACCCATGGCAACCATGAA
 ACCTCTGAAGACAGTTGCAGCATCATCAGGATTCCTTTTATCAGCATGACAAGTTTCCTC
 TTCCTCCATGTCCAATTGTTTTGATCCATAGGATCGTTCTACACTCTCTTAACATAATATA
 TATTTTACTCTATCTGATTATTGTATACAAGGATAAAATTTGATTCTTTCCTTAATGAG
 TGAGAAATATTGGAAGTGTAAAAA

>G971 Amino Acid Sequence (conserved domain in aa coordinates: 120-186)

MLDLNLKIFSSYNEDQDRKVLPMISTTGEESNSSSSSTTDSAADAFIAFGILKRDDDL
 VPPPPPPHKETGDLFPVADARRNIEFSVEDSHWLNLSLQRTQKMKVKSRRGPRSR
 SQYRGVTFYRRRTGRWESHIWDCGKQVYLGGFDYAAAAAYDRAAIKFRGLDADINRVVD
 DYRHDIDKMKNLNKVEFVQTLRRRESASFGRGSSKYKGLALQKCTQFKTHDQIHLFQNRGW
 DAAAIKYNELGKGEGAMKFGAHIKGNHNDLELSLGISSSSSEIKLTTGDYKGINRSTM
 GLYKQSSIFLPMATMKPLKTVAASSGFPFISMTSSSSSMSNCFDP*

>G975 (58..657)

ATTACTCATCATCAAGTTCTTCTCTCTGACAAACATCACAGAGTAAGTAAGAATG
 GTACAGACGAAGAAGTTTCAAGGTGTCAGGCAACGCCATTGGGGTCTTGGGTCGCTGAG
 ATTCGTCATCCTCTCTTGAACCGGAGGATTTGGCTAGGGACGTTGAGACCGCAGAGGAG
 GCAGCAAGAGCATAACGACGAGGCCCGCTTTTAAATGAGCGGCCGCAACGCCAAAACCAAC
 TTTCCCTCAACAACAACAACACCGGAGAACTTCCGAGGGCAAAACCGATATTTAGCT
 TCGTCCACAATGTCATCCTCAACATCATCTTCATCGCTCTCTTCCATCCTCAGCGCCAAA
 CTGAGGAAATGCTGCAAGTCTCCTTCCCCATCCCTCACCTGCCTCCGTCTTGACACAGCC
 AGCTCCCATATCGGCGTCTGGCAGAAACGGGCCGGTTCAAAGTCTGACTCCAGCTGGGTC
 ATGACGGTGGAGCTAGGTCCCGCAAGCTCCTCCCAAGAGACTACTAGTAAAGCTTCACAA
 GACGCTATTCTTGCTCCGACCTGAAGTTGAAATTTGGTGGCAGCAGAGAAGAAGTATTG
 GATGAGGAAGAAAAGGTTGCTTTGCAAATGATAGAGGAGCTTCTCAATACAACTAAATC
 TTATTTGCTTATATATATGTACCTATTTTCATTGCTGATTACAGCCAAAATAATCAATT
 ATACCGTGATTTTATAGATGTTTATATTAAGGTTGTTAGATATA

>G975 Amino Acid Sequence (domain in AA coordinates: 4-71)

MVQTKKFRGVRQRHWGWSWVAEIRHPLLRRIWLGTFTABEAAARAYDEAAVLMSSGRNAKT
 NFPLNNNTGETSEGKTDISASSTMSSTSSSSSLSSILSAKLKCKCKSPSPSLTCLRLDT
 ASSHIGVWQKRAGSKSDSSWMTVELGPASSSQETTSKASQDAILAPTEVEIGGSREEV
 LDEEEKVALQMIIEELLNTN*

>G994 (180..917)

TGTATATATAGTTAGTTAGTTGAGATAAACTTGGTTACCACTTTTGTGTGGTCTTTCTTT
 TTCTTTTCTCCATTTTCCATTTATCGACCCCTTGGGTGTAGCTAATTACTTTTCGCGATT
 TTCAAATCCAATAAAGTTTAAATTTGATGAAGCTTTTAAACCATATAATATAAATAA
 TGGGTGGTCGTAAACCATGTTGTGATGAGGTTGGATTAAGAAAGGTCATGGACAGTGG
 AAGAAGATGGGAAACTAGTTGATTTCTTAAGGGCACGTGGCAACTGCGGTGGTGGTGGAG
 GAGGATGGTGTGGAGAGACGTGCCAAAACCTGGCGGGGCTAAGGAGGTGTGGCAAAAGTT
 GCCGTCTCCGGTGGACTAATTATCTCCGGCCAGATCTCAAGAGAGGTCTTTTACTGAAG
 AAGAAATCCAACCTAGTCATTGATCTTCATGCTCGCTTGGCAATAGATGGTTCGAAGATTG
 CAGTGGAGTTACCAGGAAGAACAGACAACGATATCAAAAATTATTGGAACACTCATATAA
 AGAGGAAGCTTATAAGAATGGGTATTGATCCAAACACACATCGTCGATTGACCAACAAA
 AAGTCAACGAGGAGGAACGATATTGGTCAACGATCCAAAGCCTCTGTCTGAGACCGAGG
 TATCTGTTGCTTTGAAGAATGACACGTGAGCAGTGTATCAGGAAATCTAAACCAATTGG
 CTGACGTGGACGGTGATGATCAGCCGTGGAGCTTTCTAATGGAAAATGACGAAGGAGGAG
 GTGGCGACGCCCGGAGAGCTTACGATGCTATTGTCCGGTGACATTACGTCATCATGTT
 CTTCTTCGTATCTTTGTGGATGAAGTATGGAGAATTCGGATACGAAGATTTAGAATTG
 GATGTTTCGATGTTTAGAGATTCAAGTATGTTTAAATAGGCCGTAGGTTGATTAATCATA
 AGGTTCAATTGACTTCATTCTAGAAATTTGTAGTTGGACAGTATAAAGAATCAAAGTTAT
 GAAACATTGTAATTTGATTTCCAAATTAATCTAATGAATAAATGTGCTTTGCAAAAAA
 AAAAAA

>G994 Amino Acid Sequence (domain in AA coordinates: 14-123)

MGGRKPCCEVGLRKGPTVEEDGKLVDLRLARGNCGGGGGWCWRDVPKLAGLRRCGKS

CRLRWNTNYLRPDLKRGLFTEEEIQLVIDLHARLGNRWSKIAVELPGRDNDIKNYWNTHI
KRKLIRMGIDPNTHRRFDQQKVNEEETILVNDPKPLSETEVSVALKNDTSAVLSGNLNQL
ADVDDQDPWPSFLMENDEGGGGDAAGELTMLLSGDITSSSSSSSLWMKYGEFGYEDLEL
GCFDV*

>G2347 (81..626)

AGCCCATCCTTCAACATTGCTTCCTAACCAGAAATCCACCATCATCTTCCCACGAATACA
ACTTAAAGCTTTTACCAGAAAATGGAGGGTCAGAGAACACAACGCCGGGGTTACTTGAAAG
ACAAGGCTACAGTCTCCAACCTTGTGAAGAAGAAATGGAGAATGGCATGGATGGAGAAG
AGGAGGATGGAGGAGACGAAGACAAAAGGAAGAAGGTGATGGAAAGAGTTAGAGGTCCTA
GCACTGACCGTGTTCATCGCGACTGTGCCAGGTCGATAGGTGCACTGTTAATTTGACTG
AGGCCAAGCAGTATTACCGCAGACACAGAGTATGTGAAGTACATGCAAAGGCATCTGCTG
CGACTGTTGCAGGGGTCAGGCAACGCTTTTGTCAACAATGCAGCAGGTTTCATGAGCTAC
CAGAGTTTGATGAAGCTAAAAGAAGCTGCAGGAGGCGCTTAGCTGGACACAATGAGAGGA
GGAGGAAGATCTCTGGTGACAGTTTGGAGAAGGGTCAGGCCGAGAGGGTTTAGCGGTC
AAGTATCCAGACTCAAGAAAGAAAACAGGGTAGACAGGAAACTTCCTATGACCAACTCAT
CATTCAAGCGACCACAGATCAGATAAACCCCTCCCGCTCTCTCTCTCTGTCATCTACATA
TGCTCTATCTACACTCTTATTAGACAAATAATGGCATCTAACAATGTCAAGAAAAGTTGG
TCATGGTATTAAATCCTACACGGATATATAACTATAAACCTCTAGTCCCTCTATGCTGT
CTGTAATGAATATCTATCCGAAATGTATTTCGCATAGTCTTGCGTCTAATAATGTTTAT
TGATTTTGTGTA

>G2347 Amino Acid Sequence (domain in AA coordinates: 60-136)
MEGQRTQRRGYLKD KATVSNLVEEEMENGMDGEEEDGGDEDKRKKVMERVRGPSTDRVPS
RLCQVDRCTVNLTEAKQYYRRHRVCEVHAKASAATVAGVRQRFCCQCSRPFHELPEFDEAK
RSCRRRLAGHNERRRKISGDSFGE GSGRRGFSGQLIQTQERNRVDRKLPMTNSSFKRPQI
R*

>G2010 (1..525)

ATGGAGGGTAAGAGATCACAAGGACAAGGTTACATGAAAAAGAAGTCTTACCTTGTGGAA
GAAGATATGGAGACTGATACGGATGAAGAAGAGGAAGTAGGTAGGGATAGAGTTAGAGGG
TCTAGAGGTAGCATCAATCGTGGTGGCTCGTTGCGGCTTTGCCAAGTAGATAGATGCACA
GCTGATATGAAGAGGCAAACTGTATCACCAGGAGACACAAAGTGTGTGAAGTTCATGCA
AAGGCATCTTCTGTCTTTCTCTCAGGACTTAACCAACGCTTTTGTCAACAATGCAGTAGG
TTTCATGACCTCCAAGAGTTTGATGAAGCTAAGAGAAGTTGCAGGAGGCGCTTAGCTGGA
CACAATGAGCGAAGAAGGAAGAGCTCTGGTGAGAGTACTTATGGAGAAGGATCAGGTCGG
AGAGGAATCAATGGTCAGGTGGTGATGCAGAATCAAGAAAGATCAAGGGTAGAGATGACA
CTTCTATGCCAACTCATCATTCAAGCGACCACAGATTAGATAG

>G2010 Amino Acid Sequence (domain in AA coordinates: 53-127)
MEGKRSQGQGYMKKSYLVEEDMETDTEDEEEVGRDRVRGSRGSINRGGSLRLCQVDRCT
ADMKEAKLYHRRHKVCEVHAKASSVFLSGLNQRFCCQCSRFDLQEFDEAKRSCRRRLAG
HNERRRKSSGESTYGE GSGRRGINGQVVMQNQERSRVEMTLMPMNSSFKRPQIR*